



TESIS - TE142599

**IDENTIFIKASI INFORMASI PEREKONOMIAN PADA  
MEDIA SOSIAL MENGGUNAKAN METODE  
*DECISION TREE* UNTUK BADAN PUSAT  
STATISTIK**

PRAMANA YHOGA CHANDRA KUSUMA  
NRP 2214206724

DOSEN PEMBIMBING  
Dr. SURYA SUMPENO, ST., M.Sc.  
Dr. ADHI DHARMA WIBAWA, ST., MT.

PROGRAM MAGISTER  
BIDANG KEAHLIAN TELEMATIKA CIO  
JURUSAN TEKNIK ELEKTRO  
FAKULTAS TEKNOLOGI INDUSTRI  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA  
2017





TESIS - TE142599

**IDENTIFIKASI INFORMASI PEREKONOMIAN PADA  
MEDIA SOSIAL MENGGUNAKAN METODE  
*DECISION TREE* UNTUK BADAN PUSAT  
STATISTIK**

PRAMANA YHOGA CHANDRA KUSUMA  
NRP 2214206724

DOSEN PEMBIMBING  
Dr. SURYA SUMPENO, ST., M.Sc.  
Dr. ADHI DHARMA WIBAWA, ST., MT.

PROGRAM MAGISTER  
BIDANG KEAHLIAN TELEMATIKA CIO  
JURUSAN TEKNIK ELEKTRO  
FAKULTAS TEKNOLOGI INDUSTRI  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA  
2017



## LEMBAR PENGESAHAN

Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar  
Magister Teknik (MT.)

di

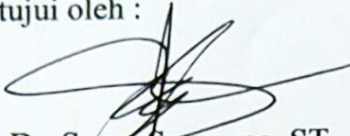
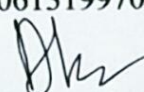

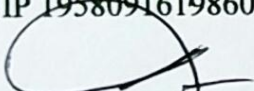
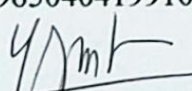
Institut Teknologi Sepuluh Nopember

Oleh:

Pramana Yhoga Chandra Kusuma  
NRP 2214206724

Tanggal Ujian : 6 Januari 2017  
Periode wisuda : Maret 2017

Disetujui oleh :

1.  Dr. Surya Sumpeno, ST., M.Sc. (Pembimbing I)  
NIP 196906131997021003
2.  Dr. Adhi Dharma Wibawa, ST., MT. (Pembimbing II)  
NIP 197605052008121003
3.  Prof. Dr. Ir. Mauridhi Hery Purnomo, M.Eng. (Penguji)  
NIP 195809161986011001
4.  Dr. Ir. Endroyono, DEA. (Penguji)  
NIP 196504041991021001
5.  Dr. Ir. Voyon Kusnendar Suprpto, M.Sc. (Penguji)  
NIP 195409251978031001



*Halaman ini sengaja dikosongkan*

## PERNYATAAN KEASLIAN TESIS

Dengan ini saya menyatakan bahwa isi keseluruhan Tesis saya dengan judul “**IDENTIFIKASI INFORMASI PEREKONOMIAN PADA MEDIA SOSIAL MENGGUNAKAN METODE *DECISION TREE* UNTUK BADAN PUSAT STATISTIK**” adalah benar-benar hasil karya intelektual mandiri, diselesaikan tanpa menggunakan bahan-bahan yang tidak diijinkan dan bukan merupakan karya pihak lain yang saya akui sebagai karya sendiri.

Semua referensi yang dikutip maupun dirujuk telah ditulis secara lengkap pada daftar pustaka. Apabila ternyata pernyataan ini tidak benar, saya bersedia menerima sanksi sesuai peraturan yang berlaku.

Surabaya, Januari 2017

Pramana Yhoga Chandra Kusuma

NRP. 2214206724

*Halaman ini sengaja dikosongkan*



# **IDENTIFIKASI INFORMASI PEREKONOMIAN PADA MEDIA SOSIAL MENGGUNAKAN METODE *DECISION TREE* UNTUK BADAN PUSAT STATISTIK**

Nama mahasiswa : Pramana Yhoga Chandra Kusuma  
NRP : 2214206724  
Pembimbing : 1. Dr. Surya Sumpeno, ST., M.Sc.  
2. Dr. Adhi Dharma Wibawa, ST., MT.

## **ABSTRAK**

Badan Pusat Statistik (BPS) adalah salah satu lembaga pemerintah yang memiliki tugas menyediakan data yang akurat dan handal bagi pemerintah dan masyarakat. Saat ini BPS belum memiliki dasar yang dapat digunakan untuk menyediakan data yang dibutuhkan oleh masyarakat. Hal tersebut bisa menyebabkan ketidaksesuaian antara data yang dihasilkan BPS dan yang dibutuhkan oleh masyarakat. BPS juga belum memperhatikan media sosial yang memiliki potensi untuk menjangkau opini terutama mengenai data yang dibutuhkan masyarakat.

Tujuan dari penelitian ini adalah mengidentifikasi topik-topik utama pada perekonomian yang bisa dijadikan dasar pemenuhan kebutuhan data kepada masyarakat. Hasil identifikasi ini diharapkan bisa menjadi saran untuk BPS dalam memaksimalkan penyediaan data kepada masyarakat.

Beberapa komentar masyarakat terutama tentang topik perekonomian pada Twitter dikumpulkan dan diklasifikasikan menggunakan metode *Decision tree*. Pengumpulan *tweet* dilakukan selama dua bulan yaitu pada bulan Januari dan Februari 2016. Proses pengumpulan menghasilkan 33.497 *tweet* dan diambil 1.858 sebagai data latih. Hasil klasifikasi menunjukkan bahwa 16.479 *tweet* termasuk topik perekonomian yang tidak menyebutkan data BPS.

Kandidat topik utama yang berhasil diidentifikasi dari *tweet* perekonomian tersebut berjumlah 28. Kandidat tersebut diidentifikasi menggunakan bigram dan trigram. Kandidat topik utama tersebut diseleksi kembali untuk menemukan topik yang relevan. Kesimpulannya, kami menemukan 3 topik utama tentang perekonomian yang perlu diperhatikan oleh BPS yaitu : paket bijak ekonomi (paket kebijakan ekonomi), sawit indonesia, dan ekonomi kreatif.

Kata kunci: **media sosial, opini publik, perekonomian, badan pusat statistik**

*Halaman ini sengaja dikosongkan*

# **IDENTIFICATION OF ECONOMIC INFORMATION ON SOCIAL MEDIA USING DECISION TREE FOR STATISTICS INDONESIA**

By : Pramana Yhoga Chandra Kusuma  
Student Identity Number : 2214206724  
Supervisor(s) : 1. Dr. Surya Sumpeno, ST., M.Sc.  
2. Dr. Adhi Dharma Wibawa, ST., MT.

## **ABSTRACT**

Badan Pusat Statistik (BPS) or known as Statistics Indonesia is a government agency which has the task of providing accurate and reliable data for the government and society. Currently, BPS does not yet have a base that can be used to provide the data needed by the community. This can lead a mismatch between the data produced by BPS and that needed by the society. BPS also has not noticed that social media has the potential to solicit opinions, especially regarding the data required by the society.

The purpose of this study was to identify the main topics on the economy that could be the basis of fulfilling the needs of data to the public. The result of this identification is expected to be a suggestion for BPS to maximize the provision of data to the public.

Some public comments, especially about topics on Twitter economies gathered and classified using Decision tree. The collection process of tweets for two months, on January and February 2016. The collecting process produce 33.497 tweets and 1.858 taken as training data. The results show that 16.479 tweets included the topic of the economy that do not mention the BPS.

Topic's candidates that successfully identified in the economic tweet totaled 28. The candidates identified using Bigram and trigram. The main topic's candidates were filtered to find the relevant topics. In conclusion, we found three main topics about the economy that need to be considered by the BPS namely: package of economic policy, Indonesian palm oil, and the creative economy.

**Key words: social media, public opinion, economics, statistics indonesia**

*Halaman ini sengaja dikosongkan*

## KATA PENGANTAR

Syukur alhamdulillah penulis panjatkan kehadiran Allah SWT yang telah memberikan limpahan karuniaNya. Selesaiannya penyusunan tesis dengan judul “Identifikasi Informasi Perekonomian Pada Media Sosial Menggunakan Metode Decision Tree Untuk Badan Pusat Statistik” ini adalah salah satu karunia terbesar yang telah diberikanNya. Tesis ini disusun untuk memenuhi salah satu persyaratan memperoleh gelar Magister Teknik (M.T.) dalam bidang keahlian Telematika CIO pada program studi Teknik Multimedia dan Jaringan Jurusan Teknik Elektro Institut Teknologi Sepuluh Nopember Surabaya.

Selesaiannya tesis ini bukanlah hasil kerja keras penulis sendiri namun banyak pihak yang telah memberikan bimbingan, kontribusi dan bantuan. Oleh karena itu pada kesempatan ini penulis menyampaikan rasa hormat dan terima kasih sebesar-besarnya kepada :

1. Bapak Dr. Surya Sumpeno, ST., M.Sc. atas ilmu, bimbingan, arahan serta waktu yang telah diberikan kepada penulis sehingga tesis ini bisa selesai. Dorongan dan motivasi yang diberikan sangat membantu penulis untuk melakukan penelitian ini.
2. Bapak Dr. Adhi Dharma Wibawa, ST., MT. atas petunjuk, bimbingannya dan bantuannya sehingga mengarahkan penulis untuk mengikuti seminar internasional ICITISEE.
3. Bapak Dr. Eko Mulyanto Yuniarno, ST., MT. atas arahannya dan waktunya selama menjadi dosen wali penulis.
4. Bapak Prof. Dr. Ir. Mauridi Hery Purnomo, M.Eng., bapak Dr. Ir. Endroyono, DEA., serta bapak Dr. Ir. Yoyon Kusnendar Suprpto, M.Sc. yang telah memberikan saran dan masukan sewaktu ujian tesis.
5. Bapak Dr. I Ketut Eddy Purnama, ST., MT. selaku kepala prodi TMJ yang telah memberikan fasilitas untuk perkuliahan serta penyelesaian tesis.

6. Seluruh dosen prodi TMJ dan Jurusan Elektro yang telah memberikan ilmu yang sangat bermanfaat untuk penyelesaian tesis. Tak lupa pula pegawai pascasarjana yang telah melayani kebutuhan penulis.
7. Orang tua penulis, bapak Tri Purnomo dan ibu Amirulsari, serta bapak Hadi Suryanto dan ibu Titik Rakhmawatien yang selalu memberikan doa, dukungan dan motivasi kepada penulis sehingga tesis ini terselesaikan. Tidak lupa juga penulis berterima kasih kepada saudara-saudara yang telah mendukung terselesaikannya tesis ini.
8. Reni Puspitasari yang selalu mendampingi, memberikan semangat, menjadi teman diskusi. Fahri dan Aisyah juga selalu memberikan semangatnya melalui segala tingkah lakunya.
9. Rekan-rekan CIO 2014 dan 2015 yang berjuang keras bersama-sama.
10. Badan Pusat Statistik yang telah memberikan kesempatan untuk melakukan studi magister ini.
11. Kementerian Koinfo yang telah menyediakan beasiswa sehingga program magister ini bisa diperoleh.
12. Semua pihak yang tidak bisa penulis sebutkan satu per satu.

Dengan keterbatasan pengalaman, ilmu maupun pustaka yang ditinjau, penulis menyadari bahwa tesis ini masih ada kekurangan dan pengembangan lanjut agar benar benar bermanfaat. Oleh sebab itu, penulis sangat mengharapkan kritik dan saran agar tesis ini lebih sempurna serta sebagai masukan bagi penulis untuk penelitian dan penulisan karya ilmiah di masa yang akan datang.

Surabaya, Januari 2017

Penulis

Pramana Yhoga Chandra Kusuma

## DAFTAR ISI

LEMBAR PENGESAHAN .....	iii
PERNYATAAN KEASLIAN TESIS .....	v
ABSTRAK .....	vii
ABSTRACT .....	ix
KATA PENGANTAR .....	xi
DAFTAR ISI .....	xiii
DAFTAR GAMBAR .....	xv
DAFTAR TABEL .....	xvii
DAFTAR RUMUS .....	xix
DAFTAR VARIABEL .....	xxi
BAB 1 PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	9
1.3 Tujuan .....	9
1.4 Batasan Masalah .....	10
1.5 Kontribusi .....	10
BAB 2 KAJIAN PUSTAKA .....	11
2.1 Ekonomi .....	11
2.2 Twitter .....	11
2.3 Twitter Archiver .....	12
2.4 Klasifikasi Manual .....	13
2.5 <i>Stopwords Removal</i> .....	14
2.6 <i>Stemming</i> .....	15
2.7 <i>Data Mining</i> .....	16
2.8 <i>Naive Bayes Classification</i> .....	18
2.9 <i>Decision Tree</i> .....	19

2.10	<i>N-gram</i> .....	23
2.11	WEKA .....	24
BAB 3 METODOLOGI PENELITIAN .....		25
3.1	Metode Penelitian.....	25
3.2	Pengumpulan <i>Tweet</i> .....	25
3.3	Klasifikasi Manual .....	27
3.4	<i>Cleaning</i> .....	30
3.5	<i>Case Folding</i> .....	30
3.6	<i>Stopwords Removal</i> .....	31
3.7	<i>Stemming</i> .....	31
3.8	Pembelajaran .....	33
3.9	Klasifikasi.....	42
3.10	Identifikasi Topik Utama.....	46
3.11	Data BPS Bidang Perekonomian.....	48
BAB 4 HASIL DAN PEMBAHASAN .....		49
4.1	Pengumpulan <i>Tweet</i> .....	49
4.2	Klasifikasi Manual .....	49
4.3	Pra Proses .....	51
4.4	Pembelajaran .....	51
4.5	Klasifikasi.....	56
4.6	Identifikasi Kandidat Topik Utama.....	58
4.7	Identifikasi Jenis Data yang Dihasilkan BPS .....	63
4.8	Seleksi Kandidat Topik Utama.....	64
BAB 5 PENUTUP .....		75
5.1	Kesimpulan.....	75
5.2	Saran.....	75
DAFTAR PUSTAKA.....		77
LAMPIRAN .....		79
BIOGRAFI .....		81



## DAFTAR GAMBAR

Gambar 1-1 Persentase Pengguna Internet Berdasarkan Pekerjaan.....	1
Gambar 1-2 Persentase Jenis Lapangan Usaha Pengguna Internet Yang Bekerja..	2
Gambar 1-3 Persentase Perangkat Yang Digunakan Untuk Mengakses Internet...	3
Gambar 1-4 Persentase Kegiatan Yang Dilakukan Oleh Pengguna Internet.....	4
Gambar 1-5 Jumlah Pengguna Twitter di Dunia Berdasarkan Negara .....	5
Gambar 1-6 Irisan Media Sosial, Media Massa Dan Kehidupan Nyata .....	6
Gambar 1-7 <i>Social Platform</i> Teraktif di Indonesia pada Januari 2016.....	7
Gambar 2-1 <i>Tweet</i> Tentang Data BPS .....	12
Gambar 2-2 Tampilan Halaman Referensi Sastrawi Stemmer .....	16
Gambar 2-3 Contoh Pohon Keputusan .....	20
Gambar 2-4 Ilustrasi Pohon Keputusan <i>Node</i> Pertama.....	22
Gambar 2-5 Ilustrasi Pohon Keputusan <i>Node</i> Kedua .....	23
Gambar 2-6 Penggunaan Metode <i>Decision Tree</i> pada WEKA.....	24
Gambar 3-1 Alur Metodologi Penelitian .....	25
Gambar 3-2 Pilihan Aturan Pencarian pada Twitter Archiver.....	26
Gambar 3-3 Hasil Pengumpulan <i>Tweet</i> oleh Twitter Archiver.....	26
Gambar 3-4 Pembagian Kelas yang akan Dibentuk .....	27
Gambar 3-5 Alur Kerja Klasifikasi Secara Manual .....	28
Gambar 3-6 Contoh Kode VBA yang Digunakan untuk <i>Stopword Removal</i> .....	31
Gambar 3-7 Kode PHP untuk Mengaktifkan <i>Library</i> Sastrawi Stemmer.....	32
Gambar 3-8 Format <i>File CSV</i> untuk Input WEKA .....	33
Gambar 3-9 Format <i>File ARFF</i> yang Kompatibel dengan WEKA .....	33
Gambar 3-10 Format Pilihan pada Proses Konversi <i>Tweet</i> ke <i>Word Vector</i> .....	34
Gambar 3-11 Ringkasan Hasil Konversi <i>Tweet</i> ke <i>Word Vector</i> .....	35
Gambar 3-12 Hasil Pembelajaran Menggunakan <i>Decision Tree</i> .....	36
Gambar 3-13 Potongan Hasil Pencarian Pohon Keputusan.....	37
Gambar 3-14 Klasifikasi Menggunakan Pohon Keputusan .....	38
Gambar 3-15 <i>Node</i> dan <i>Leaf</i> Pertama yang Terbentuk.....	41
Gambar 3-16 Pohon Keputusan yang Terbentuk .....	41
Gambar 3-17 Penentuan Kelas Menggunakan Pohon Keputusan.....	42

Gambar 3-18 Pohon Keputusan Hasil Proses di WEKA.....	42
Gambar 3-19 Penggunaan <i>FilteredClassifier</i> pada WEKA .....	43
Gambar 3-20 Format <i>Data Test</i> yang Digunakan pada WEKA.....	44
Gambar 3-21 Format <i>Data Test</i> yang Terbaca oleh WEKA .....	44
Gambar 3-22 Contoh Hasil Klasifikasi pada <i>Data Test</i> Menggunakan WEKA....	45
Gambar 3-23 Format <i>Worksheet</i> Hasil Konversi dari Klasifikasi.....	46
Gambar 3-24 Tampilan Identifikasi Unigram, Bigram dan Trigram Berdasarkan Frekuensi Kemunculan .....	47
Gambar 3-25 Kode VBA untuk Identifikasi Unigram, Bigram dan Trigram.....	47
Gambar 3-26 Tampilan Halaman Utama <i>Website</i> Badan Pusat Statistik.....	48
Gambar 4-1 Pohon Keputusan yang Berhasil Dibentuk.....	53
Gambar 4-2 Proses Klasifikasi <i>Tweet</i> pada Pohon Keputusan yang Terbentuk....	54
Gambar 4-3 Hasil Klasifikasi <i>Tweet</i> Contoh.....	55
Gambar 4-4 Alur Kerja Pengelompokan N-gram.....	60
Gambar 4-5 Pengelompokan N-gram yang Memiliki Kemiripan Makna .....	62
Gambar 4-6 Alur Kerja Seleksi Kandidat Topik Utama .....	64
Gambar 4-7 Proses Seleksi Kandidat Topik Utama .....	65
Gambar 4-8 Rilis Paket Kebijakan Ekonomi Pemerintah Indonesia.....	68
Gambar 4-9 Tren Suku Bunga Bank Indonesia dari Agustus 2013 sampai dengan Februari 2016.....	69
Gambar 4-10 Tren Harga Minyak Dunia 2012-2016 .....	70
Gambar 4-11 Tren Bursa Efek Indonesia 2005-2016.....	71
Gambar 4-12 Tren Nilai Tukar Rupiah terhadap Dolar Amerika .....	72

## DAFTAR TABEL

Tabel 2-1 Contoh Data <i>Tweet</i> yang Belum Diklasifikasikan.....	13
Tabel 2-2 Contoh Data <i>Tweet</i> yang Sudah Diklasifikasikan Secara Manual .....	14
Tabel 2-3 Beberapa <i>Stopwords</i> pada Bahasa Indonesia Menurut Tala (2003) .....	15
Tabel 2-4 Contoh Data Masukan untuk <i>Decision Tree</i> .....	19
Tabel 2-5 Contoh <i>Rule</i> yang Terbentuk dari Pohon Keputusan .....	20
Tabel 2-6 Contoh Data dan Pencarian Gain untuk Menentukan Node.....	21
Tabel 2-7 Contoh Data dan Pencarian Gain untuk Menentukan Node Kedua .....	22
Tabel 2-8 Contoh Ekstraksi Unigram, Bigram, dan Trigram dari Kalimat .....	23
Tabel 3-1 Perbedaan Format ARFF <i>Tweet</i> setelah Dikonversi ke <i>Word Vector</i> ..	36
Tabel 3-2 <i>Confussion Matrix</i> dan Akurasi Model.....	37
Tabel 3-3 Contoh <i>Dataset Tweet</i> yang Telah Diklasifikasikan Secara Manual....	39
Tabel 3-4 Konversi <i>Tweet</i> ke Atribut.....	39
Tabel 3-5 Penghitungan <i>Gain</i> tiap Atribut.....	40
Tabel 4-1 Atribut yang Berhasil Dikumpulkan oleh Twitter Archiver.....	49
Tabel 4-2 Hasil Klasifikasi <i>Tweet</i> secara Manual.....	50
Tabel 4-3 Contoh <i>Tweet</i> yang telah Dilakukan tahap Pra Proses .....	51
Tabel 4-4 Tingkat Akurasi Model yang Dibentuk Menggunakan Metode Naive Bayes .....	52
Tabel 4-5 Tingkat Akurasi Model yang dibentuk Menggunakan Metode <i>Decision Tree</i> .....	53
Tabel 4-6 Ringkasan Hasil Klasifikasi <i>Tweet</i> Menggunakan Metode <i>Decision Tree</i> .....	55
Tabel 4-7 Daftar Unigram Berdasarkan Frekuensi Kemunculan.....	59
Tabel 4-8 Daftar Bigram Berdasarkan Frekuensi Kemunculan .....	59
Tabel 4-9 Daftar Trigram Berdasarkan Frekuensi Kemunculan.....	60
Tabel 4-10 Hubungan antar N-gram dan Penentuan Kandidat Topik Utama.....	61
Tabel 4-11 Jenis Data pada Kelompok Ekonomi dan Perdagangan yang Dimiliki oleh Badan Pusat Statistik.....	63
Tabel 4-12 Hasil Seleksi Kandidat Topik Utama .....	65

*Halaman ini sengaja dikosongkan*

## DAFTAR RUMUS

Rumus 2-1 Rumus Umum Naive Bayes .....	18
Rumus 2-2 Rumus Umum <i>Entropy</i> .....	21
Rumus 2-3 Rumus Umum <i>Gain</i> .....	21

*Halaman ini sengaja dikosongkan*

## DAFTAR VARIABEL

$P$	: Nilai peluang
$A$	: Suatu kejadian A
$B$	: Suatu kejadian B
$n$	: Jumlah kejadian
$E$	: Entropy
$S$	: Jumlah kasus
$k$	: Jumlah kelas
$E$	: <i>Entropy</i>
$p_i$	: Peluang kemunculan instance pada suatu kelas
$S_i$	: Jumlah kasus pada tiap atribut

*Halaman ini sengaja dikosongkan*



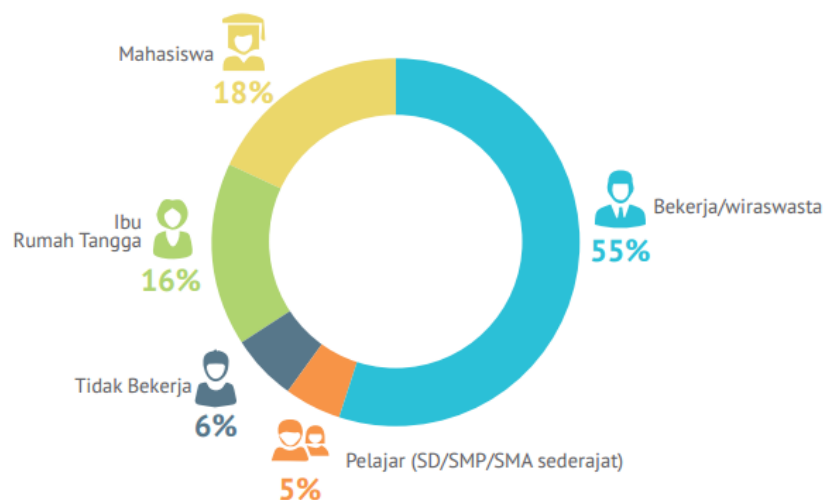
# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Perkembangan bidang Teknologi Informasi dan Komunikasi semakin meningkat dari tahun ke tahun. Teknologi yang populer di masyarakat adalah internet, terutama dengan semakin terjangkaunya harga akses internet. Salah satu organisasi penyelenggara jasa internet di Indonesia yang cukup besar adalah Asosiasi Pengguna Jasa Internet Indonesia yang biasa disingkat APJII. Berdasarkan data dari (APJII, 2015) pada tahun 2014-2015 menunjukkan bahwa jumlah pengguna internet di Indonesia 88,1 juta orang. Jumlah ini mencapai 34,9% dari total penduduk Indonesia. Pengguna internet yang cukup banyak tersebut hanya berpusat pada pulau Jawa dan Sumatera yang mencapai 70,2 juta pengguna.

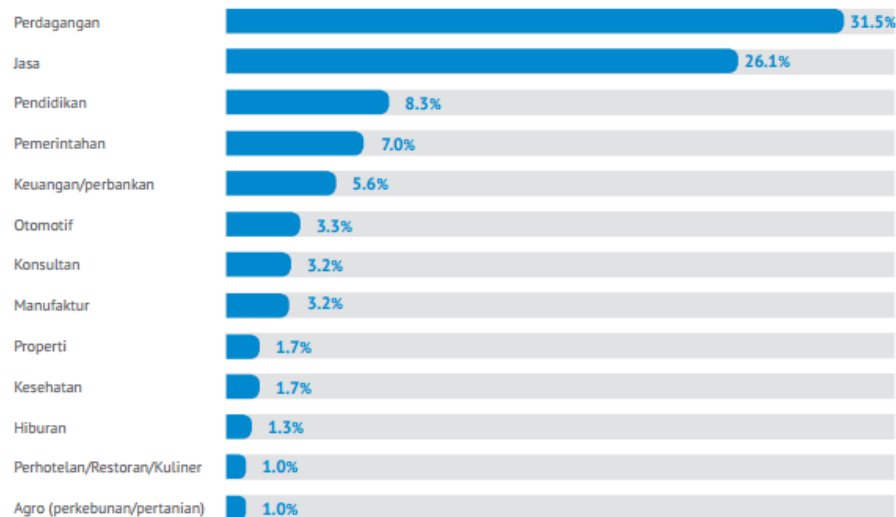
Pengguna internet di Indonesia didominasi oleh penduduk yang bekerja dengan jumlah 55%. Seperti ditunjukkan pada gambar 1-1. Dari pengguna internet yang bekerja tersebut sebanyak 57,6% bekerja pada bidang perdagangan dan jasa. Hal ini mengindikasikan bahwa kalangan pengusaha dan pelaku kegiatan ekonomi mendominasi jumlah pengguna internet di Indonesia secara keseluruhan. Dengan begitu sangat besar kemungkinan kegiatan perekonomian di Indonesia didukung oleh internet.



Gambar 1-1 Persentase Pengguna Internet Berdasarkan Pekerjaan

Sumber : Profil Pengguna Internet Indonesia 2014 (APJII, 2015)

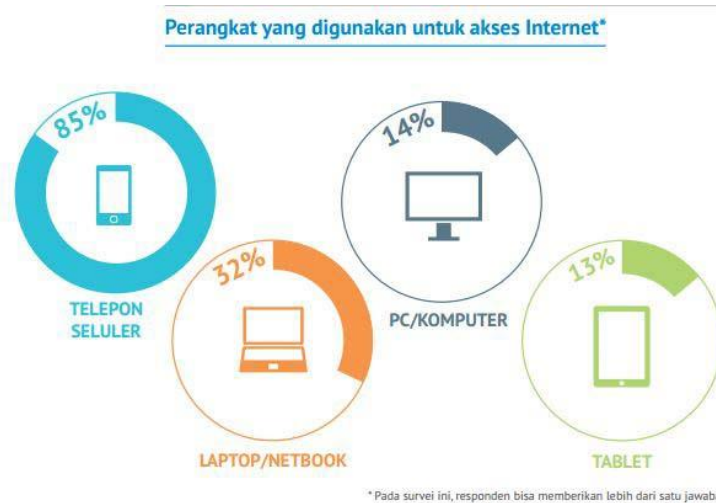
Pengguna internet di Indonesia paling banyak berusaha pada sektor perdagangan. Sektor jasa juga memiliki persentase yang cukup besar. Hal tersebut terlihat pada gambar 1-2.



Gambar 1-2 Persentase Jenis Lapangan Usaha Pengguna Internet Yang Bekerja  
Sumber : Profil Pengguna Internet Indonesia 2014 (APJII, 2015)

Indikasi besarnya pemanfaatan internet oleh pelaku ekonomi ini cukup menggembirakan karena sifat internet yang global. Internet memungkinkan komunikasi yang mudah dengan dunia. Sehingga pelaku ekonomi di Indonesia bisa berhubungan dengan sesama pelaku ekonomi di seluruh dunia. Hal itu akan sangat bermanfaat bagi pelaku ekonomi untuk mengembangkan usahanya.

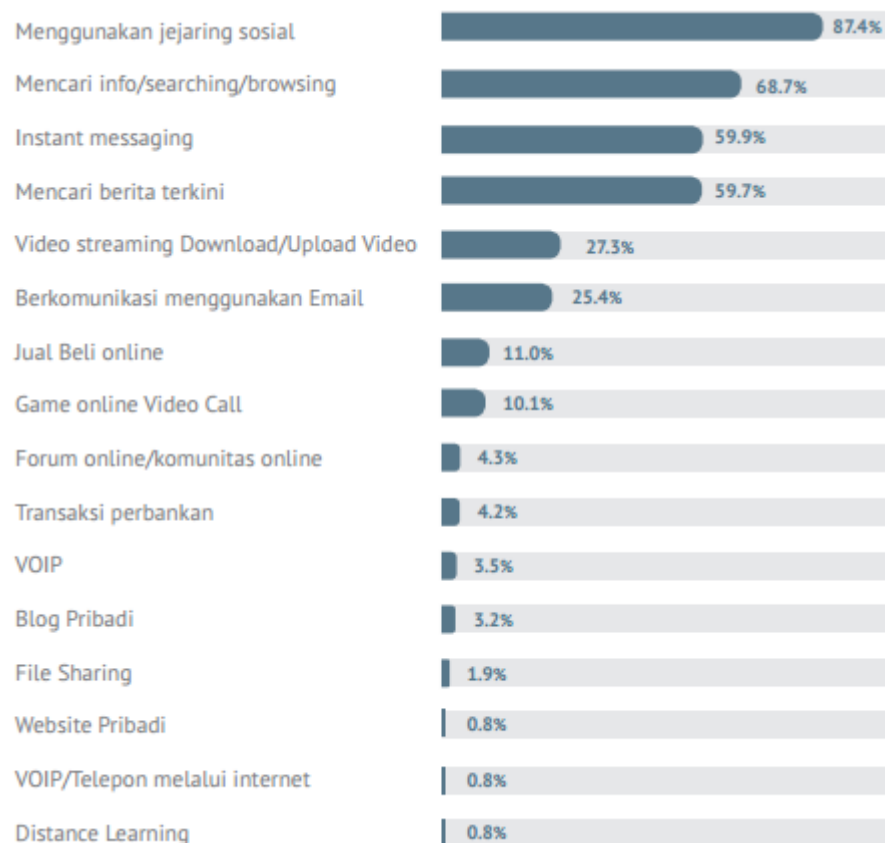
Fakta menarik lainnya dari pengguna internet di Indonesia adalah teknologi akses yang digunakan. Berdasarkan (APJII, 2015) pada Profil Pengguna Internet Indonesia 2014 diketahui bahwa 85% pengguna memanfaatkan telepon seluler dalam mengakses internet. Hal tersebut ditunjukkan pada gambar 1-3. Teknologi telepon seluler menawarkan kemudahan kepada penggunanya yaitu portable dan connected. Selain itu perkembangan yang pesat pada layanan internet di telepon seluler juga memberikan nilai tambah yang menyebabkan manusia semakin tergantung kepada telepon seluler dan internet.



Gambar 1-3 Persentase Perangkat Yang Digunakan Untuk Mengakses Internet  
 Sumber : Profil Pengguna Internet Indonesia 2014 (APJII, 2015)

Semakin berkembangnya jenis layanan yang ditawarkan internet membuat pengguna internet semakin tergantung kepadanya. Sebagai salah satu contoh yaitu pengiriman dokumen yang dahulu harus dikirimkan secara fisik ke tujuan sekarang sudah mulai digantikan oleh email. Pengiriman yang didukung oleh internet tidak hanya sebatas dokumen saja namun bisa data, suara, foto, sampai dengan multimedia seperti video. Bahkan kegiatan pribadi seperti transaksi perbankan saat ini sudah bisa dilakukan melalui internet. Internet mulai berhasil menggeser hampir sebagian besar cara manusia dalam beraktivitas.

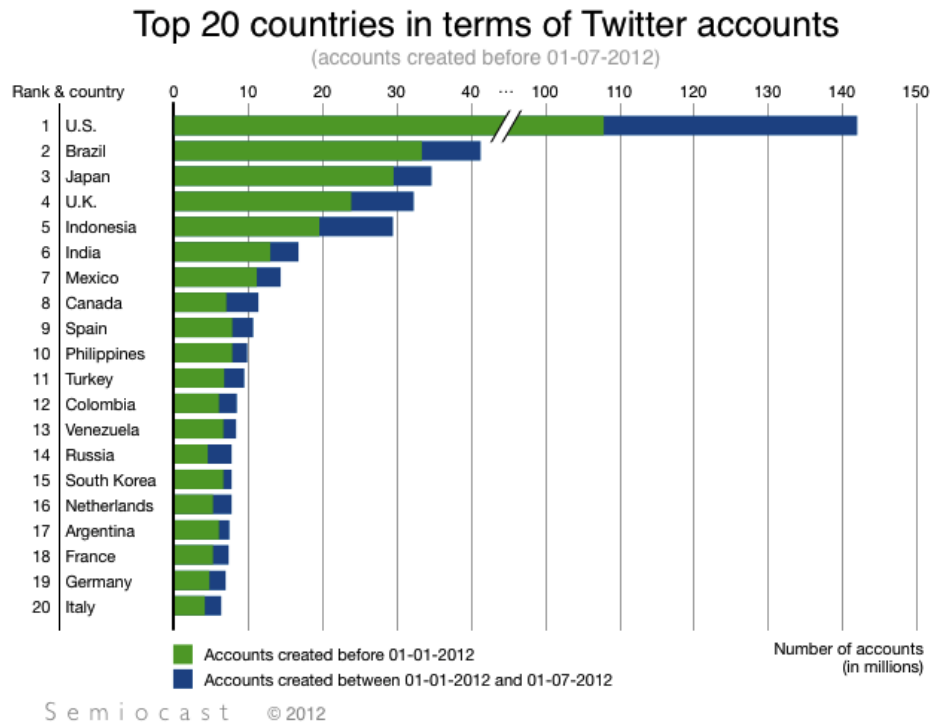
Jenis layanan internet yang bisa digunakan di Indonesia mengikuti tren perkembangan global. Jenis layanan yang paling banyak digunakan oleh penduduk Indonesia adalah akses ke media sosial. Pengguna internet Indonesia yang menggunakan internet untuk mengakses media sosial berjumlah 87,4%, disusul kemudian oleh *browsing*, *instant messaging* serta pencarian berita terkini. Jenis akses tersebut diperlihatkan pada gambar 1-4. Populernya media sosial mungkin bisa dipengaruhi oleh kondisi dan sifat sosial yang dimiliki penduduk Indonesia. Namun kemudahan akses seperti yang ditawarkan oleh telepon seluler juga sangat mungkin mempengaruhi.



Gambar 1-4 Persentase Kegiatan Yang Dilakukan Oleh Pengguna Internet

Sumber : Profil Pengguna Internet Indonesia 2014 (APJII, 2015)

Media sosial mengintegrasikan berbagai macam layanan internet seperti teks, foto bahkan video sehingga berpotensi menimbulkan ketertarikan. Tidak bisa dipungkiri bahwa penggunaan media sosial di Indonesia sangat berkembang. Salah satu *website* penilai *Key Performance Indicator* untuk media sosial menyatakan bahwa pada tahun 2012 Indonesia merupakan negara terbesar kelima dalam hal jumlah pengguna Twitter. Kota Jakarta dan Bandung juga masuk 6 besar kota di dunia yang paling sering menulis *tweet*.



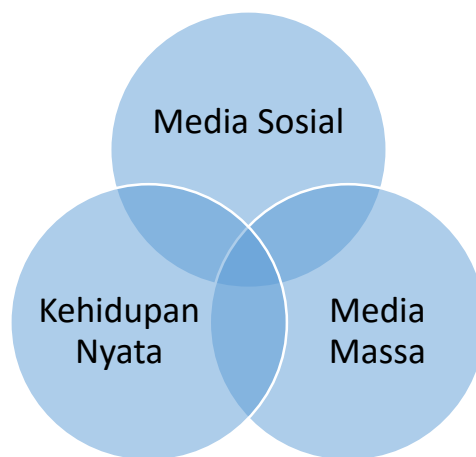
Gambar 1-5 Jumlah Pengguna Twitter di Dunia Berdasarkan Negara

Sumber : <http://semiocast.com/>

Media sosial lain yang juga populer di Indonesia adalah Facebook. Menurut (Sembiring, 2013) pada laporannya mengatakan bahwa situs media sosial yang paling banyak diakses adalah Facebook dan Twitter. Jumlah pengguna Twitter di Indonesia merupakan yang terbesar ke-5 di dunia. Hal tersebut terlihat pada gambar 1-5. Faktor kebebasan bisa berpengaruh pada perkembangan pengguna media sosial. Namun untuk memastikan hal ini masih diperlukan penelitian yang lebih mendalam.

Kepopuleran media sosial pasti disebabkan karena ada sesuatu yang menarik di dalamnya. Seperti yang disampaikan di atas bahwa kebebasan berpendapat mungkin menjadi salah satu faktor yang disukai manusia. Pada media sosial manusia bisa berinteraksi dengan sesama pengguna dari segala penjuru dunia dengan bebas tanpa terlalu mengkhawatirkan efek pembicaraannya. Pembicaraan pada media sosial sangat lengkap seperti topik bahasan di dunia nyata. Topik pembahasan bisa digolongkan dalam beberapa tema seperti perekonomian, pertanian, teknologi, berita, dan lain-lain.

Pembicaraan di media sosial sangat mungkin memiliki pengaruh ke dunia nyata. Salah satu yang mendukung hal tersebut adalah media massa dan media elektronik. Media massa dan media elektronik saat ini seringkali mengangkat topik yang sedang hangat di media sosial. Beberapa kajian telah dilakukan untuk mengetahui pengaruh media massa terhadap kehidupan nyata manusia. Menurut (Dalyono, 2010) media massa memiliki pengaruh signifikan pada pembentukan sikap dan perilaku manusia. Dengan begitu secara tidak langsung ada indikasi bahwa terdapat hubungan antara pembahasan di media sosial dan pembahasan di kehidupan nyata. Hal tersebut terlihat pada gambar 1-6. Sehingga kecenderungan pada kehidupan nyata bisa dicerminkan pada media sosial.

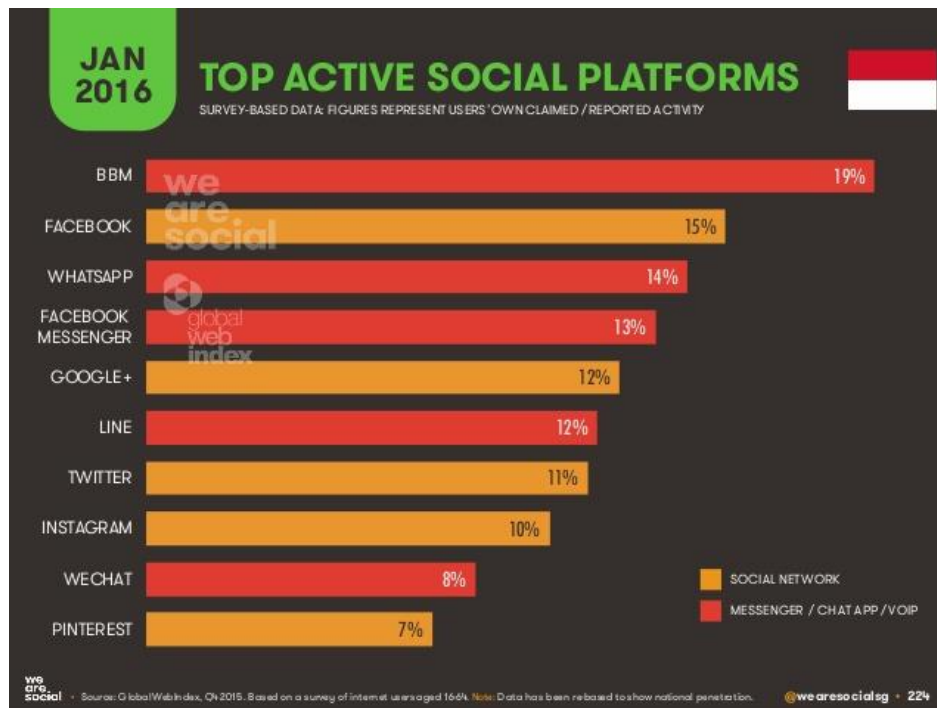


Gambar 1-6 Irisan Media Sosial, Media Massa Dan Kehidupan Nyata

Sumber : Kontribusi Konsumsi Media Massa terhadap Pemanfaatan Teknologi Informasi dan Tingkat Modernitas Generasi Muda Kota (Dalyono, 2010).

Pembahasan topik pada kehidupan nyata manusia juga bisa digolongkan menjadi beberapa topik. Pembahasan tentang perekonomian adalah pembicaraan yang seringkali selalu populer karena berhubungan langsung dengan kesejahteraan dan kebahagiaan manusia. Selain menarik perhatian masyarakat, topik perekonomian juga mendapat porsi yang besar pada pemerintahan. Presiden RI telah merumuskan program prioritas yang disebut NAWA CITA. Agenda ke enam dan tujuh dengan jelas menyebutkan program pada bidang perekonomian yaitu produktivitas dan daya saing internasional serta kemandirian ekonomi. Dua hal

tersebut jelas berhubungan dengan kehidupan masyarakat Indonesia terutama pada bidang perekonomian.



Gambar 1-7 *Social Platform* Teraktif di Indonesia pada Januari 2016

Sumber : 2016 *Digital Yearbook*

Media sosial yang paling banyak digunakan oleh masyarakat Indonesia adalah Facebook. Kemudian di urutan selanjutnya adalah twitter. Berdasarkan data dari (Kemp, 2016) pada 2016 Digital Yearbook, urutan media sosial dari yang paling banyak digunakan di Indonesia adalah BBM, Facebook, Whatsapp, Facebook Messenger, dan seterusnya. Sedangkan Twitter ada pada urutan ke 7. Tingkat keaktifan pengguna media sosial tersebut ditunjukkan pada gambar 1-7. Beberapa alasan yang diindikasikan menyebabkan populernya facebook adalah :

1. Facebook mempunyai kemampuan untuk berkirim berkas baik teks, suara, foto maupun video. Dengan kemampuan itu Facebook memberikan peluang yang besar untuk melibatkan sisi emosional pengguna.
2. Jumlah karakter yang sangat banyak pada status Facebook menyebabkan pengguna lebih bebas berekspresi. Hal ini juga bisa

menjadikan Facebook seperti sebuah pusat hiburan dimana semua hal ada disitu mulai dari barang perbelanjaan, tampilan audio-video, berita dan lain-lain.

3. Antarmuka pengguna yang lebih dinamis dan menarik sehingga pengguna betah menggunakannya.
4. Banyaknya koneksi Facebook dengan berbagai layanan online lainnya sehingga membuatnya seperti layanan online satu pintu.
5. Pengaturan privasi yang memungkinkan pengguna mengatur siapa saja yang boleh melihat statusnya.

Beberapa keunggulan Facebook tersebut menjadikannya seperti sebuah pusat hiburan pada dunia digital. Hampir seluruh hiruk pikuk kehidupan nyata terdapat juga di Facebook. Berbeda dengan Facebook, Twitter dikembangkan dengan konsep yang berbeda. Twitter diciptakan untuk lebih menjadi sebuah media informasi bukan hanya media sosial saja. Keunggulan yang dimiliki oleh Facebook tidak terdapat pada Twitter. Hal itu disebabkan karena pada dasarnya Twitter memang tidak dikembangkan untuk fasilitas-fasilitas tersebut.

Sifat Twitter yang sederhana membuat penggunanya lebih fokus dalam penggunaannya. Setelan standar privasi membuat status pengguna di Twitter bisa dilihat oleh siapapun. Adanya fasilitas *following* dan *follower* membuat pengguna mendapatkan informasi sesuai kesukaannya yang bersifat sederhana dan terfokus dengan mudah. Pada perkembangannya di Indonesia terdapat indikasi bahwa tidak seluruh kalangan masyarakat menggunakan Twitter. Fasilitas yang merupakan keunggulan Facebook menjadi saringan pengguna internet di Indonesia yang akan menggunakan Twitter. Beberapa kalangan yang mungkin memilih menggunakan Twitter adalah pengguna yang membutuhkan informasi yang sederhana dan terfokus tentang topik yang diminatinya. Antarmuka pengguna Twitter membuat kita fokus pada topik yang disukai oleh pengguna tanpa melebar kemana-mana.

Badan Pusat Statistik (BPS) merupakan salah satu instansi yang bertanggungjawab kepada presiden yang memiliki tugas menyediakan data kepada pemerintah. Beberapa manfaat data BPS adalah untuk evaluasi dan perencanaan pembangunan. Orientasi pengembangan data statistik di BPS adalah mengacu ke



kebutuhkan pemerintah. Sehingga BPS harus bisa menyediakan data yang diminta oleh pemerintah pusat. Data yang dikeluarkan oleh BPS seluruhnya digunakan oleh pemerintah. Namun pada perkembangannya BPS perlu melakukan sesuatu yang lebih terhadap data yang dihasilkan. Hal ini dipicu oleh reformasi birokrasi yang diterapkan di Indonesia. Salah satu hal yang menjadi perhatian BPS adalah kepuasan masyarakat. Menurut (BPS, 2011) kepuasan masyarakat dituangkan dalam program Quick Wins yang merupakan penjabaran dari reformasi birokrasi di BPS. Program Quick Wins merupakan program percepatan yang dilaksanakan oleh BPS untuk membangun kepercayaan masyarakat melalui penyediaan data yang benar-benar menyentuh kebutuhan pengguna data, dalam hal ini termasuk masyarakat.

## **1.2 Rumusan Masalah**

Saat ini BPS diindikasikan belum memiliki dasar yang cukup valid, reliabel dan representatif mengenai ketersediaan data yang dibutuhkan masyarakat. Ketidaksesuaian tersebut bisa menimbulkan peluang pemborosan dari segi anggaran. Salah satu cara yang bisa dilakukan oleh BPS untuk mengetahui ketersediaan data yang dibutuhkan yaitu menggunakan survei terhadap masyarakat. Dengan cakupan wilayah NKRI yang luas maka hal tersebut memerlukan biaya yang cukup besar. Menyangkut kebutuhan data masyarakat Indonesia, BPS belum memberikan perhatian kepada media sosial yang memiliki potensi cukup besar dalam usaha penjangkaran opini publik.

## **1.3 Tujuan**

Penjangkaran opini publik di media sosial merupakan salah satu alternatif yang relatif lebih murah dan mudah untuk mengetahui kebutuhan masyarakat terhadap data. Dengan dilakukannya penjangkaran opini publik ini, diharapkan topik-topik utama pada media sosial (Twitter) yang berhubungan dengan perekonomian bisa ditemukan. Topik-topik utama tersebut diharapkan bisa menjadi indikator topik-topik perekonomian yang diminati oleh masyarakat.

#### **1.4 Batasan Masalah**

Penelitian ini fokus pada media sosial Twitter yang mencakup seluruh pengguna dari Indonesia. Batasan topik pembahasan yang digunakan adalah topik perekonomian yang dibahas pada media sosial Twitter yang berupa *tweet* pengguna. Untuk melakukan penyaringan topik digunakan kata kunci awal “ekonomi”.

#### **1.5 Kontribusi**

Informasi yang diperoleh dari penelitian ini diharapkan bisa menjadi masukan kepada BPS tentang topik-topik utama perekonomian yang diminati oleh masyarakat. Dengan diketahuinya topik-topik utama tersebut maka BPS bisa melakukan evaluasi dan perencanaan terhadap strategi pengumpulan dan diseminasi data. Dengan penyediaan data yang tepat maka diharapkan tugas utama BPS bisa berjalan dengan lebih optimal.

## **BAB 2**

### **KAJIAN PUSTAKA**

#### **2.1 Ekonomi**

Kata ekonomi berasal dari dari bahasa Yunani *Oikos* dan *Nomos* yang berarti rumah tangga dan aturan. Secara sederhana ekonomi diartikan sebagai aturan-aturan untuk menjalankan rumah tangga. Mengikuti perkembangannya, rumah tangga bisa berarti perusahaan, pasar, pemerintahan maupun hubungan internasional. Pada dasarnya ekonomi tersusun dari kegiatan produksi, distribusi, dan pemakaian barang-barang serta kekayaan.

Pada zaman sekarang ekonomi memiliki hubungan yang erat dengan kesejahteraan. Ekonomi juga dekat hubungannya dengan keuangan. Ekonomi secara global bisa diartikan segala kegiatan yang dilakukan untuk memenuhi kebutuhan hidup.

Sektor-sektor ekonomi juga mengalami perkembangan. Perkembangan yang terjadi adalah dari yang dahulu hanya sekedar pemenuhan rumah tangga, keluarga, desa, kecamatan, dan seterusnya sampai dengan dunia. Sektor-sektor yang berpengaruh juga semakin bervariasi dari yang dahulu hanya menyangkut keuangan sampai dengan menyangkut hubungan internasional. Sektor pertanian, industri, pertambangan dan sebagainya semakin terkait dengan ekonomi.

#### **2.2 Twitter**

Twitter merupakan salah satu media sosial yang cukup populer digunakan. Twitter adalah media sosial yang lebih mengarah pada media informasi. Sifatnya yang sederhana membuat Twitter efektif dalam menyampaikan sebuah informasi. Twitter memiliki batasan 140 karakter pada setiap komentar atau *tweet* yang dituliskan oleh pengguna. Keterbatasan ini membuat informasi yang disampaikan oleh pengguna menjadi lebih fokus. Fasilitas *following* dan *follower* membuat pengguna bisa memilih dengan mudah informasi mengenai topik apa saja yang diminatinya. Twitter bisa diakses pada <http://twitter.com>. Contoh *tweet* yang berhubungan dengan BPS terlihat pada gambar 2.1.



Gambar 2-1 *Tweet* Tentang Data BPS

Sumber : <http://twitter.com>

Informasi *tweet* yang ada bisa digunakan untuk melakukan peninjauan opini publik. Jumlah *tweet* yang sangat banyak menyebabkan informasi yang terdapat pada *tweet* juga sangat bervariasi. Menurut (Das, et al., 2014) pencarian opini publik melalui Twitter cukup menghasilkan informasi yang menarik. Penelitian tersebut melakukan analisis mengenai sentimen positif, negatif dan netral terhadap produk Samsung Galaxy. Beberapa penelitian terkait peninjauan opini publik juga dilakukan oleh (Khoiron, 2016) yang melakukan peninjauan opini publik mengenai prioritas kebijakan publik. Menurut (Khoiron, 2016) prioritas kebijakan publik bisa diidentifikasi dengan menggunakan *tweet* di Indonesia.

### 2.3 Twitter Archiver

Twitter Archiver adalah alat yang bisa digunakan untuk mengumpulkan *tweet* sesuai dengan syarat-syarat yang kita butuhkan. Alat ini cukup memiliki pilihan saringan yang lengkap mulai dari kata kunci, pengguna twitter yang spesifik, penyaringan *hashtag*, lokasi, bahasa, dan berbagai pilihan lain. Twitter Archiver merupakan *worksheet* yang disimpan pada *cloud storage* dan memiliki *macro* yang menghubungkan dengan Twitter *APIs*. Hal ini sangat memudahkan pengguna karena hanya perlu melakukan setelan pada awal dengan penyaringan

yang dibutuhkan saja. Untuk mendapatkan *tweet* yang sesuai dengan saringan, Twitter Archiver melakukan pencarian setiap 1 jam dan menambahkan ke *worksheet*. Sehingga meminimalisir interaksi pengguna dengan Twitter Archiver, dengan kata lain sudah dilakukan secara otomatis oleh Twitter Archiver. Twitter Archiver bisa dipasang melalui *add-on* yang ada pada Chrome *Webstore*.

## 2.4 Klasifikasi Manual

Pada pembelajaran menggunakan metode supervised learning dibutuhkan data latih yang sudah terklasifikasi. Data yang sudah dikelompokkan tersebut digunakan untuk membentuk model yang akan digunakan sebagai acuan klasifikasi otomatis terhadap keseluruhan data. Jika data latih belum diklasifikasikan maka perlu dilakukan klasifikasi secara manual untuk mengkategorikan data latih. Klasifikasi secara manual harus dilakukan seobyektif mungkin untuk menjaga kualitas data latih. Petugas klasifikasi manual juga harus memiliki pengetahuan yang baik pada topik yang diteliti. Selain itu, petugas diusahakan lebih dari 1 orang untuk mengurangi subyektivitas hasil klasifikasi manual. Tabel 2.1 merupakan contoh data sebelum dan sesudah dilakukan klasifikasi manual.

Tabel 2.1 Contoh Data *Tweet* yang Belum Diklasifikasikan

No.	<i>Tweet</i>	Kelompok
1.	rendahnya pertumbuhan ekonomi indonesia triwulan dibanding tahun	?
2.	foto sensus ekonomi sensus ekonomi disosialisasikan	?
3.	jurus baru stabilkan harga daging di paket ekonomi ix	?
4.	soal teror jakarta sekjen demokrat cegah kepanikan dan ketakutan	?
5.	komoditas strategis kelapa sawit butuh kehadiran negara berupa perlindungan agar tidak dimatikan kepentingan asing	?

Tabel 2.2 Contoh Data *Tweet* yang Sudah Diklasifikasikan Secara Manual

No.	<i>Tweet</i>	Kelompok
1.	rendahnya pertumbuhan ekonomi indonesia triwulan dibanding tahun	4
2.	foto sensus ekonomi sensus ekonomi disosialisasikan	4
3.	jurus baru stabilkan harga daging di paket ekonomi ix	2
4.	soal teror jakarta sekjen demokrat cegah kepanikan dan ketakutan	1
5.	komoditas strategis kelapa sawit butuh kehadiran negara berupa perlindungan agar tidak dimatikan kepentingan asing	2

Hasil klasifikasi secara manual seperti pada tabel 2.2 yang akan digunakan sebagai dasar pembuatan model untuk melakukan klasifikasi pada data sejenis yang belum diklasifikasikan. Data perlu dilakukan klasifikasi menggunakan model karena data yang dimiliki sangat banyak sehingga akan sulit melakukan klasifikasi manual terhadap data tersebut.

## 2.5 *Stopwords Removal*

Salah satu yang perlu dibersihkan dari teks adalah *stopwords*. *Stopwords* adalah daftar kata-kata yang tidak memiliki arti signifikan pada sebuah teks namun sering muncul. Penghapusan *stopwords* pada *text mining* diperlukan untuk menghilangkan fitur yang tidak penting, dalam hal ini adalah kata yang tidak memiliki arti penting. Dengan fitur yang semakin sedikit maka diharapkan proses pembentukan model lebih efisien dan tepat. Daftar *stopwords* berbeda pada setiap bahasa. Untuk bahasa indonesia *stopwords* yang bisa digunakan adalah hasil penelitian dari (Tala, 2003). Daftar tersebut berisi 758 *stopwords* dalam bahasa indonesia. Beberapa daftar *stopwords* yang digunakan ditampilkan pada tabel 2.3.

Tabel 2.3 Beberapa *Stopwords* pada Bahasa Indonesia

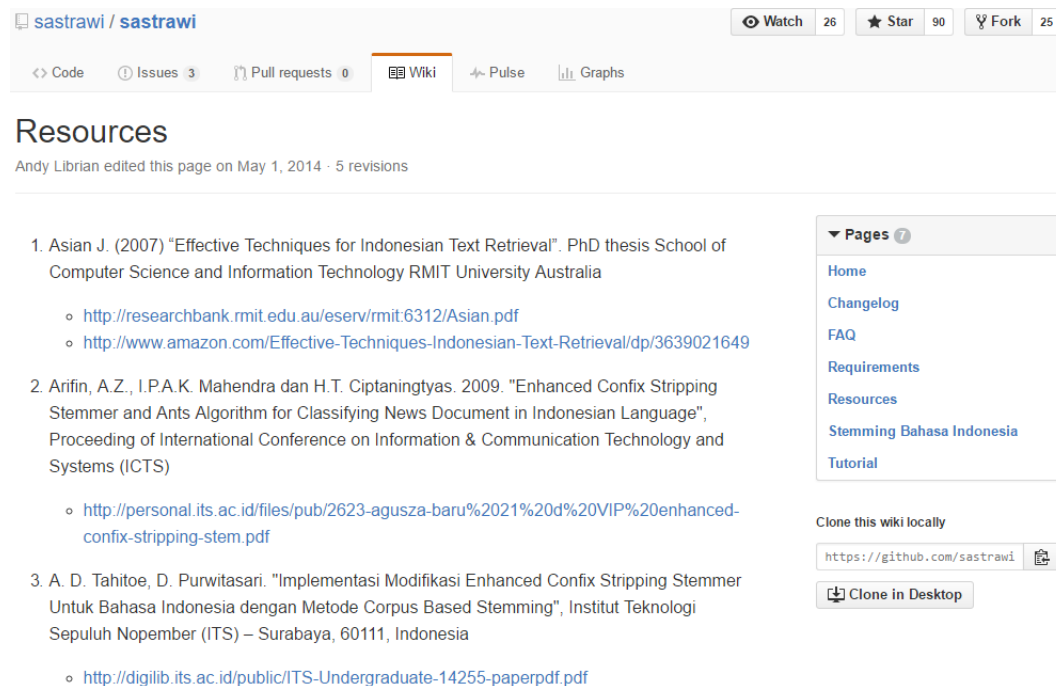
ada	Dikira	lagi	pernah	Siap
adalah	dilakukan	lagian	persoalan	siapa
adanya	dilalui	lah	pertama	siapakah
adapun	dilihat	lain	pertama-tama	siapapun
agak	dimaksud	lainnya	pertanyaan	sini
agaknya	dimaksudkan	lalu	pertanyakan	sinilah
agar	dimaksudkannya	lama	pihak	soal
akan	dimaksudnya	lamanya	pihaknya	soalnya
akankah	diminta	lanjut	pukul	suatu
akhir	dimintai	lanjutnya	pula	sudah
akhiri	dimisalkan	lebih	pun	sudahkah
akhirnya	dimulai	lewat	punya	sudahlah
aku	dimulailah	lima	rasa	supaya
akulah	dimulainya	luar	rasanya	tadi
amat	dimungkinkan	macam	rata	tadinya
amatlah	dini	maka	rupanya	tahu
anda	dipastikan	makanya	saat	tahun
andalah	diperbuat	makin	saatnya	tak
antar	diperbuatnya	malah	saja	tambah
antara	dipergunakan	malahan	sajalah	tambahnya
antaranya	diperkirakan	mampu	saling	tampak
apa	diperlihatkan	mampukah	sama	tampaknya
apaan	diperlukan	mana	sama- sama	tandas

Daftar *stopwords* ini adalah daftar kata yang akan dihilangkan dari *tweet* untuk mengurangi fitur/kata yang tidak penting. Tahap ini biasanya dilakukan pada pra proses sebelum dilakukan pembelajaran.

## 2.6 Stemming

*Stemming* adalah proses pencarian kata dasar pada sebuah kata. Proses *stemming* berbeda-beda pada tiap bahasa karena mempunyai karakteristik imbuhan yang berbeda pula. *Library* Sastrawi Stemmer merupakan salah satu alat yang bisa digunakan untuk melakukan *stemming* pada bahasa Indonesia. Sastrawi Stemmer menggunakan beberapa kombinasi algoritma untuk mendapatkan hasil yang cukup akurat. *Library* Sastrawi Stemmer bisa dijalankan melalui bahasa pemrograman PHP dan *database* MySQL. Sastrawi Stemmer bisa diakses di

<http://sastrawi.github.io/>. Contoh tampilan Sastrawi Stemmer seperti pada gambar 2.2.



Gambar 2-2 Tampilan Halaman Referensi Sastrawi Stemmer

## 2.7 Data Mining

*Data mining* adalah sebuah proses pencarian secara otomatis informasi yang berguna dalam tempat penyimpanan data berukuran besar. Istilah lain yang sering digunakan diantaranya *knowledge discovery (mining) in databases* (KDD), *knowledge extraction*, *data/pattern analysis*, *data archeology*, *data dredging*, *information harvesting*, dan *business intelligence*. Teknik *data mining* digunakan untuk memeriksa basis data berukuran besar sebagai cara untuk menemukan pola yang baru dan berguna. Tidak semua pekerjaan pencarian informasi dinyatakan sebagai *data mining*. Sebagai contoh, pencarian *record individual* menggunakan *database management system* atau pencarian halaman *web* tertentu melalui kueri ke semua *search engine* adalah pekerjaan pencarian informasi yang erat kaitannya dengan *information retrieval*. *Data mining* adalah salah satu metode yang biasa digunakan pada *machine learning*.

Tugas-tugas dalam *data mining* secara umum dibagi ke dalam dua kategori utama yaitu :



- Prediktif. Tujuan dari tugas prediktif adalah untuk memprediksi nilai dari atribut tertentu berdasarkan pada nilai dari atribut-atribut lain. Atribut yang diprediksi umumnya dikenal sebagai target atau variabel tak bebas, sedangkan atribut-atribut yang digunakan untuk membuat prediksi dikenal sebagai explanatory atau variabel bebas.
- Deskriptif. Tujuan dari tugas deskriptif adalah untuk menurunkan pola-pola (korelasi, *trend*, *cluster*, trayektori, dan anomali) yang meringkas hubungan yang pokok dalam data. Tugas data mining deskriptif sering merupakan penyelidikan dan seringkali memerlukan teknik *postprocessing* untuk validasi dan penjelasan hasil.

*Data Mining* bisa digunakan untuk beberapa tujuan yaitu :

1. Analisis asosiasi yaitu pencarian aturan-aturan asosiasi yang menunjukkan kondisi-kondisi nilai atribut yang sering terjadi bersama-sama dalam sekumpulan data. Analisis asosiasi sering digunakan untuk menganalisa market basket dan data transaksi.
2. Klasifikasi yaitu proses menemukan model (fungsi) yang menjelaskan dan membedakan kelas-kelas atau konsep, dengan tujuan agar model yang diperoleh dapat digunakan untuk memprediksikan kelas atau objek yang memiliki label kelas tidak diketahui. Model yang turunan didasarkan pada analisis dari training data (yaitu obyek data yang memiliki label kelas yang diketahui). Model yang diturunkan dapat direpresentasikan dalam berbagai bentuk seperti aturan IF-THEN klasifikasi, pohon keputusan, formula matematika atau jaringan syaraf tiruan. Dalam banyak kasus, pengguna ingin memprediksikan nilai-nilai data yang tidak tersedia atau hilang (bukan label dari kelas). Dalam kasus ini biasanya nilai data yang akan diprediksi merupakan data *numeric*. Kasus ini seringkali dirujuk sebagai prediksi. Di samping itu, prediksi lebih menekankan pada identifikasi *trend* dari distribusi berdasarkan pada data yang tersedia.
3. Analisis *cluster* yaitu melakukan analisis terhadap objek data dimana label kelas tidak diketahui. *Clustering* dapat digunakan untuk

menentukan label kelas tidak diketahui dengan cara mengelompokkan data untuk membentuk kelas baru. Sebagai contoh *clustering* rumah untuk menemukan pola distribusinya. Prinsip dalam *clustering* adalah memaksimumkan kemiripan *intra*class dan meminimumkan kemiripan *inter*class.

4. Analisis *Outlier* yaitu melakukan analisis terhadap objek data yang tidak mengikuti perilaku umum dari data. *Outlier* dapat dianggap sebagai *noise* atau pengecualian. Analisis data *outlier* dinamakan *outlier mining*. Teknik ini berguna dalam *fraud detection* dan *rare events analysis*.
5. Analisis *trend* dan *evolusi* yaitu menjelaskan dan memodelkan *trend* dari obyek yang memiliki perilaku yang berubah setiap waktu. Teknik ini dapat meliputi karakterisasi, diskriminasi, asosiasi, klasifikasi, atau *clustering* dari data yang berkaitan dengan waktu

## 2.8 Naive Bayes Classification

Salah satu pemanfaatan *Data Mining* adalah melakukan klasifikasi terhadap data, yaitu mengelompokkan (mengklasifikasikan) data ke dalam satu atau beberapa kelas yang sudah didefinisikan sebelumnya. Salah satu metode dalam klasifikasi data adalah Naïve Bayes Classifier (NBC). Naïve Bayes Classifier merupakan salah satu metode *machine learning* yang memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya.

Dasar dari Naïve Bayes yang dipakai adalah rumus Bayes :

$$P(A|B_1 \dots B_n) = \frac{P(A)P(B_1 \dots B_n|A)}{P(B_1 \dots B_n)} \quad (2.1)$$

$P(A|B_1 \dots B_n)$  = Peluang terjadinya kejadian A ketika kondisi B

$P(A)$  = Peluang terjadinya kejadian A

$P(B_1 \dots B_n|A)$  = Peluang terjadinya kejadian B ketika kondisi A

$P(B) = \text{Peluang terjadinya kejadian } B$

Cara kerja Naïve Bayes Classifier melalui dua tahapan, yaitu :

1. *Learning* (pembelajaran). Pada tahap ini diperlukan masukan berupa data yang sudah diklasifikasikan untuk membentuk model.
2. *Classify* (Pengklasifikasi). Pada tahap ini dilakukan pengklasifikasian data berdasarkan model yang sudah terbentuk pada pembelajaran.

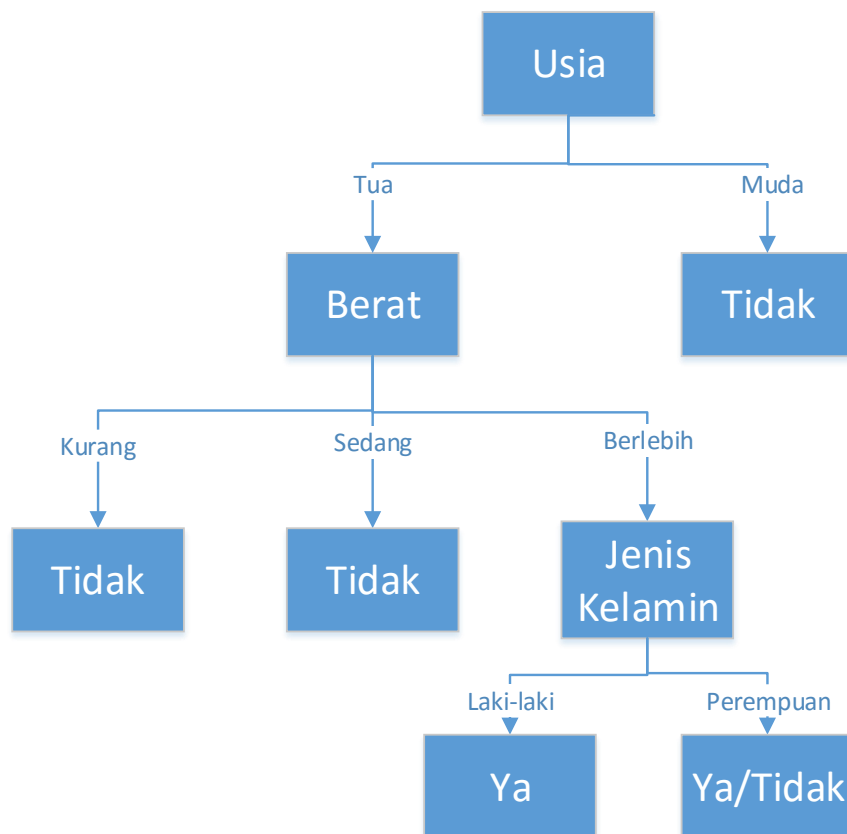
## 2.9 Decision Tree

*Decision tree* adalah salah satu metode klasifikasi dengan *supervised learning*. Pembelajaran pada *Decision tree* menggunakan data latih yang telah diklasifikasikan terlebih dahulu. Konsep *Decision tree* adalah mengubah data menjadi pohon keputusan dan selanjutnya dijadikan rule. Data adalah input yang berupa tabel yang berisi *instance* dan atribut. Salah satu atribut adalah merupakan kelas/kategori itu sendiri. Contoh data seperti pada tabel 2.4.

Tabel 2.4 Contoh Data Masukan untuk *Decision Tree*.

<b><i>Instance</i></b>	<b>Usia</b>	<b>Berat</b>	<b>Jenis kelamin</b>	<b>Hipertensi (kategori)</b>
Ani	Muda	Sedang	Wanita	Tidak
Badu	Tua	Berlebih	Pria	Tidak
Cecep	Tua	Berlebih	Pria	Ya

Data masukan tersebut diubah menjadi pohon keputusan supaya lebih mudah dipahami. Contoh pohon keputusan yang dihasilkan seperti gambar 2.3.



Gambar 2-3 Contoh Pohon Keputusan

Dari sebuah pohon keputusan kemudian dijadikan *rule* yang bisa digunakan untuk memprediksi data yang belum diklasifikasi. Contoh *rule* yang bisa dibentuk dari *decision tree* seperti pada tabel 2.5.

Tabel 2.5 Contoh *Rule* yang Terbentuk dari Pohon Keputusan

Rule1	: IF usia=muda THEN hipertensi=tidak
Rule2	: IF usia=tua ^ berat=underweight THEN hipertensi=tidak
Rule3	: IF usia=tua ^ berat=overweight ^ jeniskelamin=laki-laki THEN hipertensi=ya

Pada data biasanya terdiri lebih dari 1 atribut. Pertanyaan yang muncul pada *Decision tree* adalah bagaimana cara memilih atribut yang menjadi *node* utama dan mana yang selanjutnya. Untuk memilih *node* utama tersebut digunakan nilai *gain* tiap atribut. Atribut yang memiliki *gain* paling sedikit akan menjadi *node*

utama. Pemilihan atribut selanjutnya yang menjadi *node* selanjutnya juga menggunakan nilai *gain*. Atribut yang memiliki nilai entropy lebih sedikit menjadi *leaf* utama terlebih dahulu.

Entropy memiliki rumus umum :

$$Entropy(S) = \sum_{j=1}^k -p_i \log_2 p_i \quad (2.2)$$

$S$  = himpunan (dataset) kasus

$k$  = jumlah kelas

$p_i$  = probabilitas yang didapat dari peluang kemunculan instance

Setelah ditemukan nilai *entropy* untuk tiap atribut maka dicari nilai *gain*.

Nilai *gain* terbesar dijadikan sebagai *node*.

Rumus untuk mencari *gain* adalah :

$$Gain(T) = Entropy(S) - \sum_{i=1}^k \frac{|S_i|}{|S|} \times Entropy(S) \quad (2.3)$$

$T$  = atribut

$k$  = jumlah kelas

$S$  = himpunan (dataset) kasus

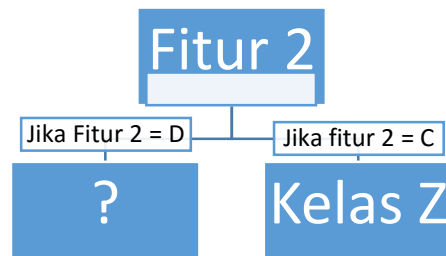
$S_i$  = himpunan (dataset) untuk pada tiap atribut

Untuk memahami proses pembentukan *decision tree* lebih mudah menggunakan contoh. Contoh penghitungan untuk menentukan *node* dan *leaf* bisa dilihat pada tabel 2.6.

Tabel 2.6 Contoh Data dan Pencarian Gain untuk Menentukan Node

<i>Instance</i>	Fitur 1	Fitur 2	Fitur 3	Fitur 4	Fitur $S_i$	Kelas
1	A	D	E	H	I	Y
2	B	D	E	G	J	Y
3	B	D	F	G	I	Z
4	A	C	E	G	J	Z
n	...	...	...	...	...	...
<i>Entropy</i>	<i>Entropy</i> $S_1$	<i>Entropy</i> $S_2$	<i>Entropy</i> $S_3$	<i>Entropy</i> $S_4$	<i>Entropy</i> $S_5$	<i>Entropy</i> $S_i$
<i>Gain</i>	<i>Gain</i> (1)	<i>Gain</i> (2)	<i>Gain</i> (3)	<i>Gain</i> (4)	<i>Gain</i> (5)	<i>Gain</i> (i)

Dari tabel 2.5. bisa diketahui nilai *Gain* untuk masing-masing fitur/atribut. Untuk menentukan *node* pertama yaitu dicari atribut yang memiliki *gain* paling tinggi. Atribut tersebut kemudian dijadikan *node* pertama. Jika dimisalkan fitur/atribut 2 memiliki *gain* paling besar maka *node* pertama terlihat seperti gambar 2-4.



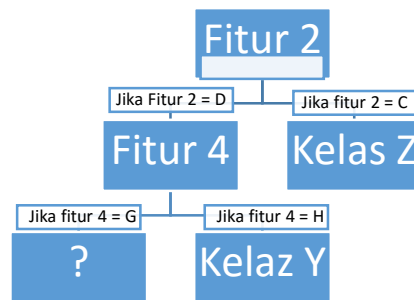
Gambar 2-4 Ilustrasi Pohon Keputusan *Node* Pertama

Untuk menentukan *node* selanjutnya (jika nilai fitur 2 = D) tetap digunakan penghitungan *gain*. Namun pada penghitungan *gain* ini data disaring untuk menghilangkan fitur 2 dan *instance* yang bernilai C karena fitur 2 dengan nilai tersebut sudah menjadi *leaf* pertama yaitu ditentukan sebagai kelas Z.

Tabel 2.7 Contoh Data dan Pencarian Gain untuk Menentukan Node Kedua

<i>Instance Number</i>	Fitur 1	Fitur 3	Fitur 4	Fitur $S_i$	Kelas
1	A	E	H	I	Y
2	B	E	G	J	Y
3	B	F	G	I	Z
n	...	...	...	...	...
<i>Entropy</i>	<i>Entropy</i> $S_1$	<i>Entropy</i> $S_3$	<i>Entropy</i> $S_4$	<i>Entropy</i> $S_i$	
<i>Gain</i>	<i>Gain (1)</i>	<i>Gain (3)</i>	<i>Gain (4)</i>	<i>Gain (5)</i>	<i>Gain (i)</i>

Misalkan pada pencarian *node* kedua pada tabel 2.7 ditemukan *gain* tertinggi ada pada fitur 4 maka ilustrasi bentuk pohon keputusan seperti pada gambar 2-5.



Gambar 2-5 Ilustrasi Pohon Keputusan *Node* Kedua

Pada gambar 2-6 terlihat sudah terbentuk pohon keputusan dengan 2 *node* dan 2 *leaf*. Proses menentukan *node* dan *leaf* selanjutnya sama dengan proses pencarian *node* dan *leaf* kedua. Proses tersebut diulang sampai menemukan fitur terakhir yang menjadi *node*. Setelah semua fitur menjadi *node* maka bentuk pohon keputusan telah selesai. Pohon keputusan tersebut digunakan untuk melakukan klasifikasi terhadap data yang belum di klasifikasikan. Pohon keputusan ini diharapkan memudahkan pengklasifikasian dengan melihat nilai fitur yang ada pada masing-masing *instance*.

## 2.10 N-gram

N-gram adalah potongan sejumlah n obyek yang bersebelahan dalam sebuah string sesuai dengan jumlah potongan yang dibutuhkan. String yang dimaksud bisa berupa kata, kalimat, maupun teks. Jumlah potongan bisa berjumlah 1, 2, 3, 4, dan seterusnya disesuaikan dengan kebutuhan. Potongan sejumlah n obyek ini bisa digunakan untuk menemukan pola obyek sehingga memiliki arti yang spesifik dan relevan terhadap topik. Contoh penentuan n-gram dari sebuah kalimat adalah seperti pada tabel 2-8.

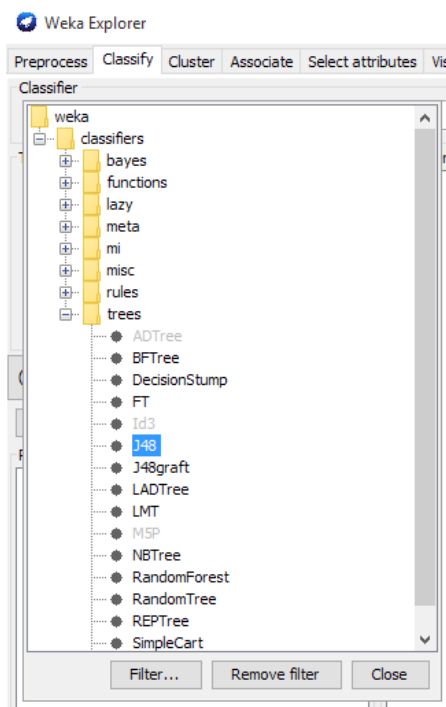
Tabel 2.8 Contoh Ekstraksi Unigram, Bigram, dan Trigram dari Kalimat

Kalimat	Unigram	Bigram	Trigram
Penulisan ilmiah memerlukan ketelitian dan kecermatan	Penulisan; Ilmiah; Memerlukan; Ketelitian; Dan; Kecermatan.	Penulisan ilmiah; Ilmiah memerlukan; Memerlukan ketelitian; Ketelitian dan; Dan kecermatan.	Penulisan ilmiah memerlukan; Ilmiah memerlukan ketelitian; Memerlukan ketelitian dan; Ketelitian dan kecermatan

## 2.11 WEKA

Penelitian ini menggunakan piranti lunak WEKA yang merupakan piranti gratis. WEKA adalah singkatan dari Waikato Environment for Knowledge Analysis. Piranti ini dikembangkan oleh Universitas Waikato di Selandia Baru berbasis Java. Piranti lunak ini adalah alat bantu dalam melakukan *Machine Learning*. Fasilitas yang tersedia cukup lengkap mulai dari alat visualisasi dan algoritma untuk analisis data serta pembangunan model untuk melakukan prediksi. Piranti ini relatif mudah digunakan karena memiliki antarmuka pengguna untuk memudahkan penggunaannya.

WEKA bisa digunakan untuk proses *data mining* mulai dari pra proses, *clustering*, klasifikasi, regresi, visualisasi, sampai dengan seleksi fitur. Metode yang didukung oleh WEKA cukup bervariasi mulai dari Bayesian, *Trees*, *Rule*, *Regresi*, SVM, dan lainnya. Kelengkapan fitur dan antarmuka yang relatif mudah menjadi alasan penulis menggunakan piranti ini. *Decision Tree* pada WEKA disebut sebagai J48. Hal tersebut seperti tampak pada gambar 2-6.



Gambar 2-6 Penggunaan Metode *Decision Tree* pada WEKA

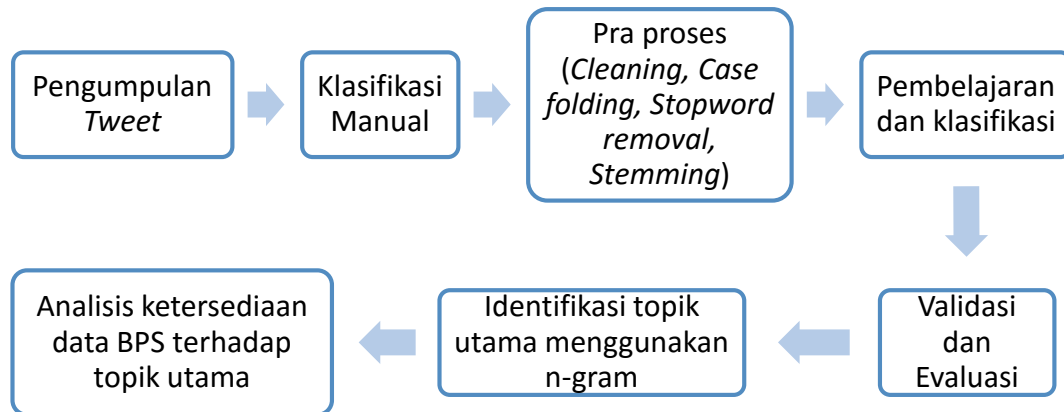


## BAB 3

### METODOLOGI PENELITIAN

#### 3.1 Metode Penelitian

Pada penelitian ini alur kerja yang digunakan seperti pada gambar 3-1.



Gambar 3-1 Alur Metodologi Penelitian

#### 3.2 Pengumpulan *Tweet*

Proses pertama pada penelitian ini adalah mengumpulkan *tweet*. *Tweet* yang dibutuhkan adalah seluruh *tweet* yang berasal dari masyarakat Indonesia. Kata kunci yang digunakan sebagai penyaring adalah “ekonomi”. Alat yang digunakan untuk mengumpulkan *tweet* adalah *Twitter Archiver*. Alat ini memanfaatkan *Twitter API* sehingga pengguna harus sudah terdaftar pada pengembang Twitter. Alat ini memiliki beberapa keunggulan yaitu :

1. Tersedia versi gratis.
2. Berjalan pada *cloud drive*, sehingga bisa berjalan secara otomatis.
3. Pilihan *filter* yang tersedia cukup banyak.

Cukup banyaknya saringan yang tersedia di *Twitter Archiver* membuat alat tersebut fleksibel digunakan. Beberapa saringan yang tersedia pada alat ini terlihat pada gambar 3-2.

Gambar 3-2 Pilihan Aturan Pencarian pada Twitter Archiver

Aturan pencarian tersebut perlu diubah untuk mendapatkan *tweet* yang berhubungan dengan perekonomian. Kata “ekonomi” diisikan sebagai kata kunci sehingga membuat Twitter Archiver menyimpan seluruh *tweet* yang mengandung kata tersebut. Proses pencarian oleh twitter archiver ini tidak hanya pada kata-kata di *tweet* saja namun juga *hyperlink* yang ada pada *tweet* juga. Twitter archiver melakukan pencarian setiap 1 jam untuk mencari *tweet* yang sesuai dengan aturan pencarian yang telah ditetapkan. Hasil pencarian akan disimpan pada sheet yang telah ada pada *cloud drive*. Kelemahan penggunaan *cloud drive* adalah terbatasnya jumlah *cell* yang bisa disimpan. Pada *cloud drive* yang digunakan pada penelitian ini jumlah *cell* yang bisa disimpan paling banyak berjumlah 2 juta. Hasil pengumpulan *tweet* terlihat pada gambar 3-3.

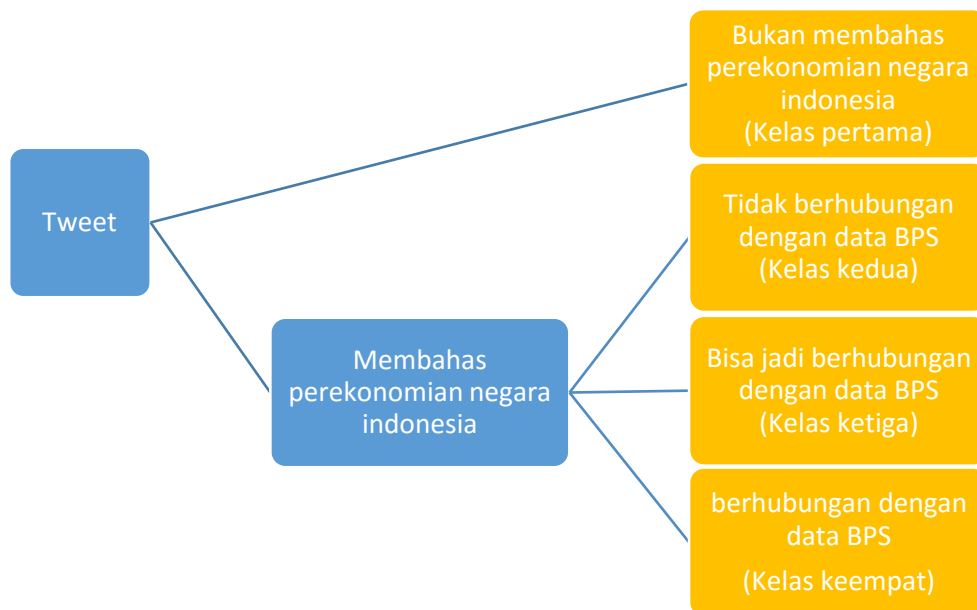
Ekonomi_v1			
File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive			
twitterfeed			
Twitter Search Rule: ekonomi lang:id			
Date	reen Nan	Full Name	Tweet Text
1/11/2016	@DAUDF	Daud ba	Ketua MPR Ungkap Tantangan Umat Islam: Tantangan itu: sumber daya manus
1/11/2016	@Hizkia_	Kiting	Ketua MPR Ungkap Tantangan Umat Islam: Tantangan itu: sumber daya manus
1/11/2016	@inpojok	Info Jabar	Ada Tol Cipali, Ekonomi Indramayu 2015 Melambat - https://t.co/w2dSFGED7I
1/11/2016	@human	Humster N	Ketua MPR Ungkap Tantangan Umat Islam: Tantangan itu: sumber daya manus
1/11/2016	@ara_bo	Cik Lemon RT @Muhammad_Baik:	kami sebagai mahasiswa terkesan dgn keadaan ekono
1/11/2016	@Nanda	Winandar	Ketua MPR Ungkap Tantangan Umat Islam: Tantangan itu: sumber daya manus
1/11/2016	@Kaskus	Kaskus Ber	Menteri Susi Bawa Kasus Hai Fa ke Pengadilan Internasional https://t.co/l4ZJJ0
1/11/2016	@Kaskus	Kaskus Ber	Menteri Susi Sebut Telah Pulangkan 23 Korban perbudakan https://t.co/2V1Lk
1/11/2016	@Kaskus	Kaskus Ber	DJP Gandeng Kemenkominfo Incar Pajak E-Commerce https://t.co/jluSprWo5x
1/11/2016	@Kaskus	Kaskus Ber	Pemprov DKI Mulai Lelang Proyek Pembangunan Infrastruktur https://t.co/YJF
1/11/2016	@alfama	Alfa Mand	Ketua MPR Ungkap Tantangan Umat Islam: Tantangan itu: sumber daya manus

Gambar 3-3 Hasil Pengumpulan *Tweet* oleh Twitter Archiver

Hasil pengumpulan *tweet* menampilkan beberapa atribut diantaranya tanggal *tweet*, nama pengguna, isi, jumlah *retweet*, jumlah *follower*, dan lain sebagainya. Hasil pengumpulan ini bisa diunduh dalam bentuk *worksheet*, *portable document*, *web page*, maupun *comma-separated values*.

### 3.3 Klasifikasi Manual

Klasifikasi manual diperlukan untuk membuat data training yang digunakan untuk proses pembelajaran. Proses ini dilakukan secara manual karena *tweet* yang diperoleh belum terbagi menjadi kelas-kelas yang diharapkan. Proses ini berperan penting karena akan mempengaruhi keseluruhan proses selanjutnya. Pembentukan kelas-kelas disesuaikan dengan kebutuhan penelitian. Pada penelitian ini dibentuk 4 kelas seperti pada gambar 3-4.



Gambar 3-4 Pembagian Kelas yang akan Dibentuk

Proses pengklasifikasi manual ini dilakukan sebelum pra proses lain untuk menjaga keaslian *tweet* tersebut. *Tweet* yang masih lengkap akan lebih mudah dipahami maksudnya. Untuk memberikan hasil yang lebih obyektif, maka pengklasifikasi manual ini dilakukan oleh 2 orang yang berstatus sebagai pegawai BPS. Pegawai BPS dipilih karena dianggap memiliki pengetahuan yang lebih baik mengenai jenis-jenis data yang dihasilkan BPS. Pemilihan *tweet* yang

akan diklasifikasikan secara manual dilakukan secara acak. Hal ini dikarenakan hasil pengumpulan *tweet* tersebut belum diketahui sama sekali pengelompokannya. Sehingga cara pemilihan secara acak dianggap menjadi cara terbaik. Alur kerja klasifikasi manual yang dilakukan seperti pada gambar 3-5.



Gambar 3-5 Alur Kerja Klasifikasi Secara Manual

Pada proses ini ditentukan 4 kelas yang akan digunakan yaitu :

1. *Tweet* yang tidak termasuk topik perekonomian negara Indonesia. *Tweet* yang digolongkan pada kelas ini diantaranya adalah *tweet* yang menggunakan bahasa Malaysia, *check in*, serta beberapa *tweet* yang mengandung kata ekonomi namun tidak berhubungan dengan perekonomian negara Indonesia. Beberapa contoh *tweet* yang masuk pada kelas ini adalah :
  - melemah nya ekonomi china berdampak kredit macet di china juga akan merembet ke negara lain dan semakin mengancam pertumbuhan ekonomi global
  - at sekolah tinggi ilmu ekonomi Indonesia stiesia
  - fokus pertumbuhan ekonomi jaksa Agung tunda eksekusi mati
  - cikgu ada berita sedih kelas ekonomi dibatalkan classmates yeaaaayyyy parrryyy cikgu cikgu je sedih kamu semua happy
  - hai sang ekonomi kreatif q msh tetap vokalist rockstar dan main alat musik q tdk mahir amat
2. *Tweet* yang memiliki topik perekonomian negara namun tidak menyinggung/menggunakan data BPS. *Tweet* yang berada di kelas ini bisa saja menyebut sebuah data yang bukan berasal dari BPS.

Selain itu, *tweet* yang memiliki topik perekonomian namun tidak memerlukan data juga masuk pada kelas ini. Beberapa contoh *tweet* yang dimasukkan pada kelas ini:

- otoritas pasar modal teken kontrak jptm dengan link bei bersama pt kliring penjaminan efek
- wakil presiden jufus kalla menginginkan berbagai pihak membayar pajak dengan jujur agar target intrins
- pertamina tertarik jajak blok masela pemerintah dalam hal ini kementerian energi dan sumber daya mineral es
- hipmi paket ekonomi pemerintah jangan terlalu banyak himpunan pengusaha muda indonesia
- profesi yang akan bersaing di program masyarakat ekonomi asean mea berhati hatilah para kaum profesional

3. *Tweet* yang memiliki topik perekonomian negara dan menyinggung/menggunakan data yang mungkin bersumber dari BPS. Kelas ini dibentuk untuk memfasilitasi *tweet* yang menyebutkan sebuah data namun tidak terlalu spesifik mengenai sumbernya. Sumber data tersebut bisa berasal dari BPS maupun di luar BPS. Beberapa contoh *tweet* yang ada pada kelas ini adalah :

- mau untung jual rumah murah pelemahan ekonomi membuat penjualan rumah anjlok dan meleset dari target yang di
- ekonomi di indonesia nurun enggak juga tuh buktinya penjualan honda jazz nyaingin honda brio
- menteri rizal ramli blok masela harus dorong ekonomi indonesia timur
- agus marto yakin ekonomi ri membaik ini alasannya
- jd peringatkan kesenjangan ekonomi sudah lampu kuning
- ekonomi membaik bank mandiri yakin kredit konsumen meningkat

4. *Tweet* yang memiliki topik perekonomian negara dan menyinggung/menggunakan data yang bersumber dari BPS. Kelas

ini berisi *tweet* yang mengandung data yang bersumber dari BPS.

Beberapa contoh *tweet* yang ada pada kelas ini adalah :

- pertumbuhan ekonomi indonesia tahun lalu persen
- bi konservatif terlalu takut mngorbankan pertumbuhan ekonomi
- percepat pertumbuhan ekonomi menkeu harap alumni lpdp jadi wiraswasta
- terkerek ekspor desember capai usd miliar ekspor indonesia tercatat naik sebesar persen dengan
- inflasi februari diprediksi terimbas kenaikan harga beras

### 3.4 *Cleaning*

Tahap ini dilakukan untuk membersihkan *tweet* dari karakter yang tidak diperlukan. Karakter yang dibersihkan antara lain *web link* (diawali “http://”), *mention* (diawali @), *hashtag* (#), *retweet* (RT), angka, serta karakter non alfabet. Tahap *cleaning* ini dilakukan pada Microsoft Excel menggunakan fasilitas *Find & Replace*. Pada beberapa pencarian diperlukan kode *Regular Expression*.

### 3.5 *Case Folding*

Tahap ini dilakukan setelah *tweet* bersih dari karakter yang tidak diperlukan. Tahap ini membuat seluruh huruf *tweet* menjadi huruf kecil. Tahap ini diperlukan untuk menyeragamkan kata-kata *tweet* sehingga fitur yang akan diseleksi lebih sedikit. Sebagai contoh pada 2 *tweet* :

1. Ekonomi melambat warga china pindahkan uang ke luar negeri
2. China tingkatkan kerja sama ekonomi dengan Iran

Jika tanpa melakukan *case folding* maka fitur yang dihasilkan dari 2 *tweet* tersebut adalah :

- |             |              |                |             |
|-------------|--------------|----------------|-------------|
| 1. Ekonomi  | 5. pindahkan | 9. negeri      | 13. sama    |
| 2. melambat | 6. uang      | 10. China      | 14. ekonomi |
| 3. warga    | 7. ke        | 11. Tingkatkan | 15. dengan  |
| 4. china    | 8. luar      | 12. kerja      | 16. Iran    |

Jika menerapkan case folding maka fitur yang terbentuk adalah :

- |             |              |                |            |
|-------------|--------------|----------------|------------|
| 1. ekonomi  | 5. pindahkan | 9. negeri      | 13. dengan |
| 2. melambat | 6. uang      | 10. tingkatkan | 14. iran   |
| 3. warga    | 7. ke        | 11. kerja      |            |
| 4. china    | 8. luar      | 12. sama       |            |

Berdasarkan contoh terlihat bahwa kata Ekonomi dan ekonomi adalah fitur yang berbeda jika tanpa *case folding*. Namun dengan *case folding* maka dua fitur tersebut menjadi sama.

### 3.6 Stopwords Removal

*Stopwords* adalah kata umum yang seringkali muncul dalam jumlah banyak dan dianggap tidak memiliki arti. Contoh *stopword* pada bahasa indonesia diantaranya ada, apa, bagi, ke, di, dari, pak, yang, dan lain sebagainya. Daftar *stopwords* yang digunakan berjumlah 758 kata (Tala, 2003). Proses ini menggunakan Microsoft Excel dan VBA. Kode VBA yang digunakan untuk proses ini seperti pada gambar 3-6.

---

#### Option Explicit

---

```
Sub test()  
    Dim a, i As Long, myPtn As String  
    With Cells(1).CurrentRegion  
        a = .Columns(1).Value  
        With .Columns(2).Offset(1).Resize(.Rows.Count - 1)  
            myPtn = Join(Application.Transpose(.Cells), Chr(2))  
        End With  
        With CreateObject("VBScript.RegExp")  
            .Global = True  
            .IgnoreCase = True  
            .Pattern = "[\\$\\(\\)\\-\\^\\|\\[\\]\\+\\*\\.\\?\\|\\]"  
            'myPtn = Replace(.Replace(myPtn, "\\$1"), Chr(2), "|")  
            .Pattern = "\\b(" & Replace(.Replace(myPtn, "\\$1"), Chr(2), "|") & ")\\b"  
            For i = 2 To UBound(a, 1)  
                a(i, 1) = Application.Trim(.Replace(a(i, 1), ""))  
            Next  
        End With  
        .Columns(4).Value = a  
    End With  
End Sub
```

Gambar 3-6 Contoh Kode VBA yang Digunakan untuk *Stopword Removal*

### 3.7 Stemming

Proses ini digunakan untuk mencari kata dasar dari sebuah kata. Pencarian ini diperlukan untuk menyeragamkan kata yang mempunyai arti sama. Proses ini memanfaatkan *library* Sastrawi yang berjalan di server PHP. Sastrawi merupakan

salah satu alat *stemmer* untuk bahasa indonesia. *Library* sastrawi ditanamkan pada server lokal PHP agar proses lebih mudah dan cepat. Kode PHP yang digunakan untuk mengaktifkan *library* Sastrawi adalah seperti gambar 3-7.

```
$query = "SELECT cleaned_stop_word, Ran_2 FROM `tweet77854`";
$result_query = $MySQLiconn->query($query);

$stemmerFactory = new \Sastrawi\Stemmer\StemmerFactory();
$stemmer = $stemmerFactory->createStemmer();

if ($result_query->num_rows > 0) {
    while($row = $result_query->fetch_assoc()) {
        echo $row['Ran_2']. ". ". $row['cleaned_stop_word']. " =====> ";

        $sentence = $row['cleaned_stop_word'];
        $output = $stemmer->stem($sentence);

        echo $output . "====>";

        $query_stemmed = "UPDATE tweet77854 SET stemmed = '". $output. "' WHERE Ran_2 = ".$row['Ran_2'];
        echo $query_stemmed. "<br>";
        $result_query_stemmed = $MySQLiconn->query($query_stemmed);
    }
}
```

Gambar 3-7 Kode PHP untuk Mengaktifkan *Library* Sastrawi Stemmer

Proses *cleaning*, *case folding*, *stopwords removal* dan *stemming* disebut juga dengan *feature selection*. Setiap kata pada *text mining* akan dijadikan sebuah fitur. Oleh karena itu diperlukan pemilihan kata yang berguna saja sehingga memperkecil fitur yang akan terbentuk. *Cleaning* digunakan untuk menghilangkan karakter yang tidak relevan dengan penelitian. *Case folding* digunakan untuk menyeragamkan kata sehingga kata dengan arti yang sama hanya tercatat sekali. *Stopword removal* digunakan untuk mengurangi fitur yang tidak mempunyai arti signifikan terhadap teks. Sedangkan *stemming* digunakan untuk menyamakan dua atau kata yang mempunyai arti sama. Sebagai contoh adalah kata “mengolah”. Pada satu kumpulan teks ada kemungkinan kemunculan kata tersebut dalam beberapa bentuk misalnya mengolah, diolah, olah, dan pengolahan. Jika kata-kata tersebut tidak diseragamkan maka akan terbentuk 4 fitur. Namun jika dilakukan stemming terlebih dahulu maka akan terbentuk 1 fitur yaitu “olah”.



### 3.8 Pembelajaran

Proses pembelajaran dilakukan setelah pra proses selesai. Pada proses pembelajaran dilakukan beberapa ujicoba menggunakan metode *decision tree* dan naive bayes. Percobaan yang dilakukan menggunakan software WEKA. *Tweet* yang telah siap di proses disimpan menjadi file *comma-separated values* untuk memudahkan konversi ke WEKA. Format file CSV biasanya dipisahkan oleh tanda koma (,) seperti terlihat pada gambar 3-8.

```
1 kategori2,stemmed_2
2 1,smg cepat kau hancur rusia jahat krisis ekonomi warga rusia antri makan gratis
3 1,dei diklat ekonomi islam dei salah kaderisasi ksei mahasiswa cont
4 1,kronologis duga palsu flight approval airfast
5 1,tuntut jihad msykt negara jihad bangun ekonomi politik alam sihat adil bebas pimpin
6 1,bikin hastag prayjkt macam ntar obrol dunia investor kurang rupiah ribu eko
7 1,gara gara gawat ekonomi ayam goreng inti angin jual pasar
8 1,perkasa ekonomi bumiputera fokus modal insan
9 1,culi bunuh anak depok duga motif ekonomi
10 1,tumbuh ekonomi cina lamban
11 1,follow sumber berita basi basi olahraga sehat ekonomi
12 1,kuliah hujan fakultas ekonomi unnm pic
13 1,perintah negara mana dunia hantam uang ekonomi cina komunis pkc embargo ekpoimport
14 1,dar krisis ekonomi payah laa ikan bilis ekor
15 1,data tumbuh ekonomi china pasar rilis data tumbuh pdb china kuartal
16 1,ekonomi perintah rencana
17 1,polres pamekasan bentuk tim khusus narkoba warta ekonomi
18 1,ktp slip gaji aju kredit card bank mega jasa
19 1,dlm agama ade jenis najis najis mukhafafah najis mutawasitah najis mughalazzah ekonomi
20 1,tahniah politik moga allah kati juang martabat ekonomi duduk bangsa melayu mata dunia
21 1,imf pangkas proyeksi tumbuh ekonomi global
22 1,kelas kah pasar fakultas ekonomi bisnis unlam
23 1,cuit bareskrim panjang tahan ongen direktorat tindak pidana ekonomi khus
24 1,lemah nya ekonomi china dampak kredit macet china rembet negara ancam tumbuh ekonomi global
25 1,with tengku muhammad yahya anggi and others fakultas ekonomi umsu
```

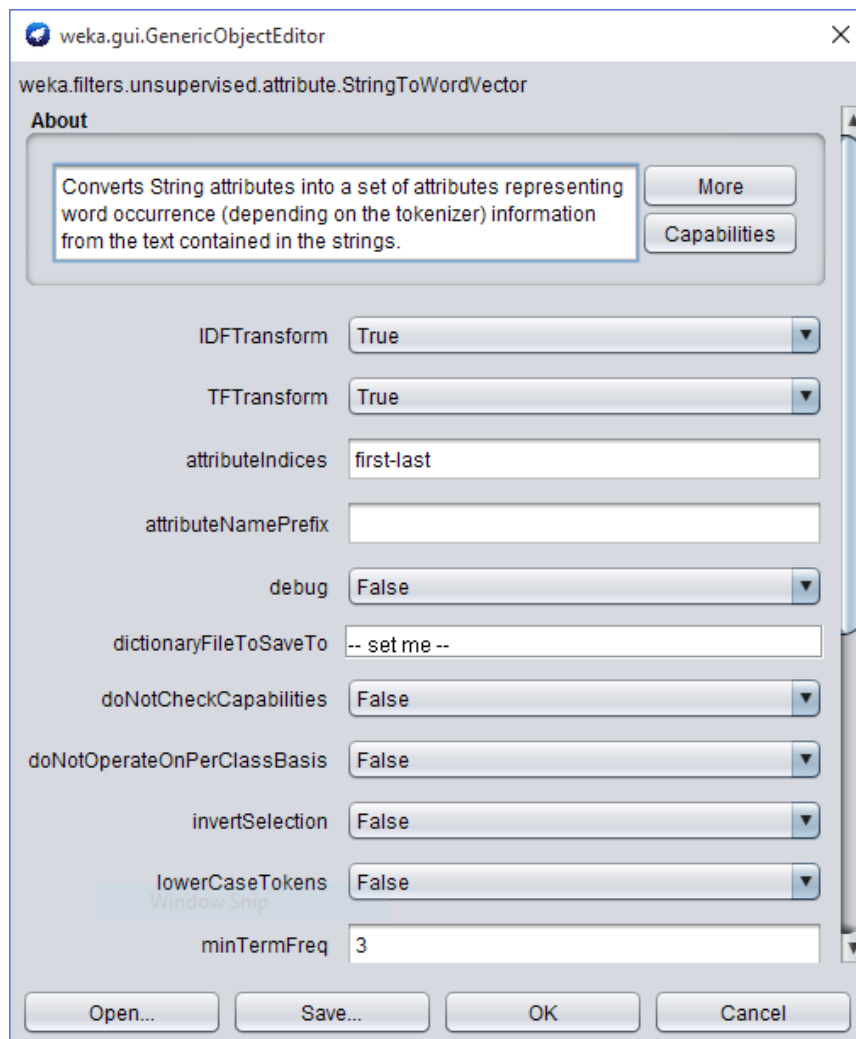
Gambar 3-8 Format *File CSV* untuk Input WEKA

File *comma-separated values* tersebut kemudian dibuka di WEKA dan disimpan sebagai file default WEKA yaitu ARFF. File ARFF memiliki bentuk yang berbeda dengan CSV. Format file ARFF terlihat pada gambar 3-9.

```
1 @relation 'Data Training
2 weka.filters.unsupervised.attribute.Reorder-R2,1-weka.filters.unsupervised.instance.Randomize-S42'
3 @attribute stemmed string
4 @attribute kategori {1,2,3,4}
5
6 @data
7 'ekonomijakarta presiden joko widodo jokowi laku',1
8 'parah pikir gpp cenah rupiah gosong prcuma idup ekonomi mkin puruk smuanya gosong',1
9 'mimpi tuju indonesia demokrasi kuat stabil matang ekonomi kuat lanjut adil perada',2
10 'analisis layak kereta cepat aspek pasar finansial manajemen sosial ekonomi teknis awat',2
11 'kredit masalah mandiri ekonomi global pengaruh makro ekonomi indonesia',1
12 'potong belanja bantu target tumbuh ekonomi',4
13 'hatta rajasa news bank indonesia ekonomi indonesia baik tribunnewsbi ekonomi indones',3
14 'tempat indonesia peringkat mudah usaha',2
15 'gatrane news bank indonesia turun giro wajib minimum dana tambah bank rupiah triliun',2
16 'perintah cepat jual bumn perintah cepat jual usaha badan usaha milik',2
17 'hai sang ekonomi kreatif msh vokalist rockstar main alat musik tdk mahir',1
18 'orasi ilmiah sby kan ekonomi indonesia presiden enam indonesia susilo yudhoyono membeberk',1
19 'tolak ihsg rupiah pagi buka anjlok persen',2
20 'klaim capai program juta rumah klaim',2
21 'lambat jua tunggu ekonomi kukuh haha',1
22 'masyarakat blm mapan hrs tumbuh lbh cepat drpd masyarakat mapan gerak ekonomi tum bpk',2
23 'rilis utara indonesia tumbuh tinggi',4
24 'perintah optimistis dpr rampung ruu ampun pajak',2
25 'agus marto ekonomi tumbuh inflasi via',4
```

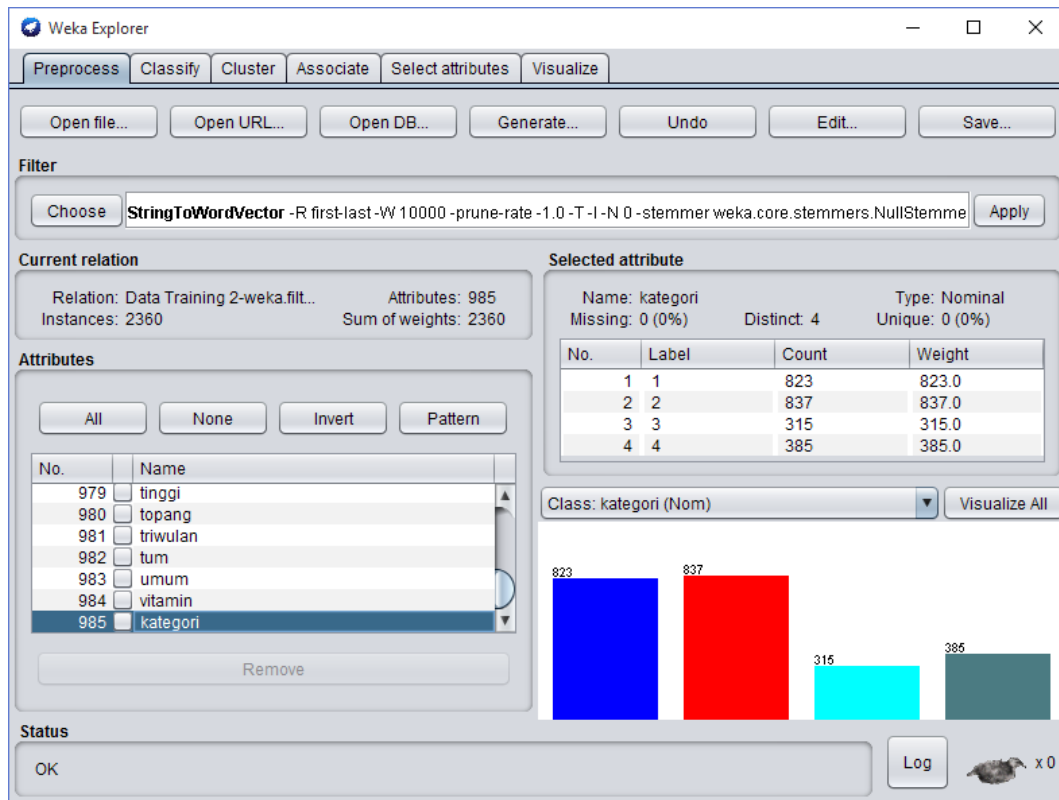
Gambar 3-9 Format *File ARFF* yang Kompatibel dengan WEKA

Pada proses penyimpanan sebagai file ARFF ini dilakukan juga proses *Randomize* untuk menata urutan *tweet* secara random. Pada pembelajaran ini digunakan *filter* StringToWordVector pada WEKA. *Filter* tersebut berfungsi untuk memisah kalimat menjadi per kata yang mana akan dijadikan sebagai fitur yang bisa diproses oleh WEKA. Pada proses ini diperlukan beberapa setelan seperti penggunaan TF dan IDF, jumlah minimal kemunculan kata, jumlah kata yang disimpan, *stemmer*, dan lain-lain. Pada penelitian ini digunakan pembobotan TF dan IDF dan jumlah minimal kemunculan kata yaitu 5. Pilihan lain seperti *stemmer*, *stopwords*, dan *tokenizer* tidak digunakan karena sebelumnya *tweet* sudah dibersihkan dari hal tersebut. Pilihan-pilihan yang tersedia tersebut terlihat pada gambar 3-10.



Gambar 3-10 Format Pilihan pada Proses Konversi *Tweet* ke *Word Vector*

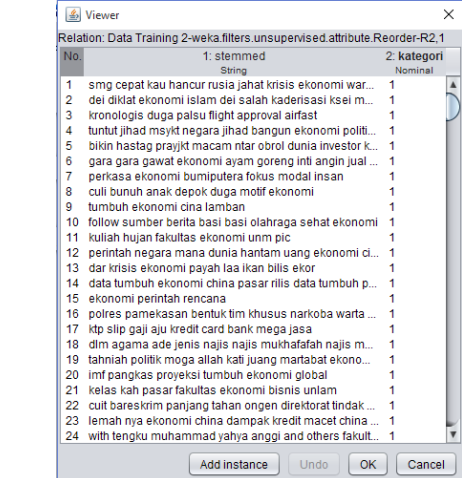
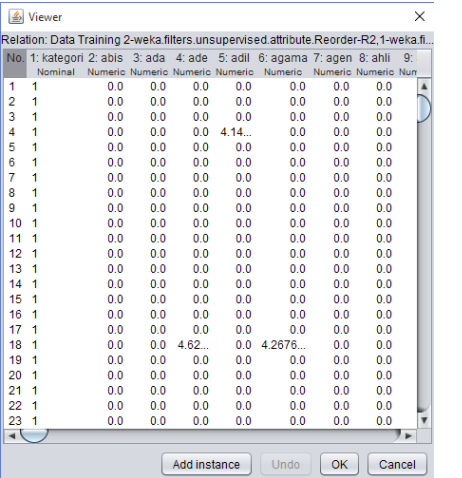
Ringkasan hasil proses konversi *tweet* ke *word vector* akan tampil pada halaman awal WEKA. Ringkasan yang muncul diantaranya jumlah fitur, jumlah data total, jumlah data tiap kelas, dan grafik. Tampilan halaman depan seperti pada gambar 3-11.



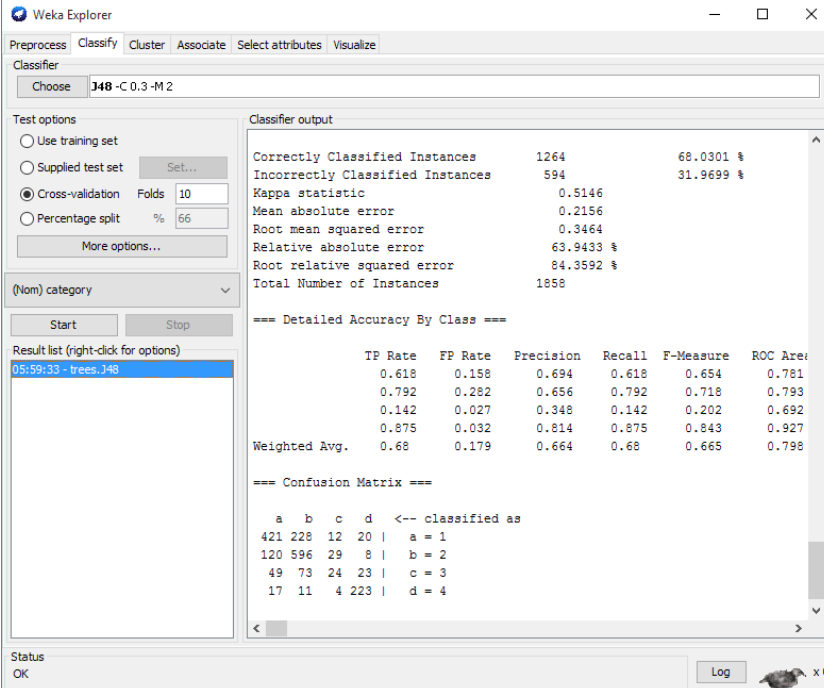
Gambar 3-11 Ringkasan Hasil Konversi *Tweet* ke *Word Vector*

Data hasil konversi ke *word vector* memiliki format yang berbeda dengan sebelumnya. Hal tersebut dikarenakan konversi telah merubah *tweet* yang sebelumnya berupa kalimat menjadi kata yang berbentuk vektor. Perbedaan format data bisa dilihat pada tabel 3.1.

Tabel 3.1 Perbedaan Format ARFF *Tweet* setelah Dikonversi ke *Word Vector*

Sebelum konversi		Setelah konversi	
			

Setelah proses konversi ke *word vector* selesai maka *data training* siap untuk diproses ke tahap selanjutnya. *Data training* tersebut dilatih dengan metode *Decision tree* dan Naive Bayes. Validasi yang digunakan adalah *cross-validation* menggunakan 10 *folds*. Gambar 3-12 menunjukkan contoh hasil pembelajaran pada data latih menggunakan metode *Decision tree*.



Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 -C 0.3 -M 2

Test options

☐ Use training set

☐ Supplied test set Set...

☒ Cross-validation Folds 10

☐ Percentage split % 66

More options...

(Nom) category

Start Stop

Result list (right-click for options)

05:59:33 - trees.J48

Classifier output

Correctly Classified Instances 1264 68.0301 %

Incorrectly Classified Instances 594 31.9699 %

Kappa statistic 0.5146

Mean absolute error 0.2156

Root mean squared error 0.3464

Relative absolute error 63.9433 %

Root relative squared error 84.3592 %

Total Number of Instances 1858

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
a	0.618	0.158	0.694	0.618	0.654	0.781
b	0.792	0.282	0.656	0.792	0.718	0.793
c	0.142	0.027	0.348	0.142	0.202	0.692
d	0.875	0.032	0.814	0.875	0.843	0.927
Weighted Avg.	0.68	0.179	0.664	0.68	0.665	0.798

=== Confusion Matrix ===

a	b	c	d	<-- classified as
421	228	12	20	a = 1
120	596	29	8	b = 2
49	73	24	23	c = 3
17	11	4	223	d = 4

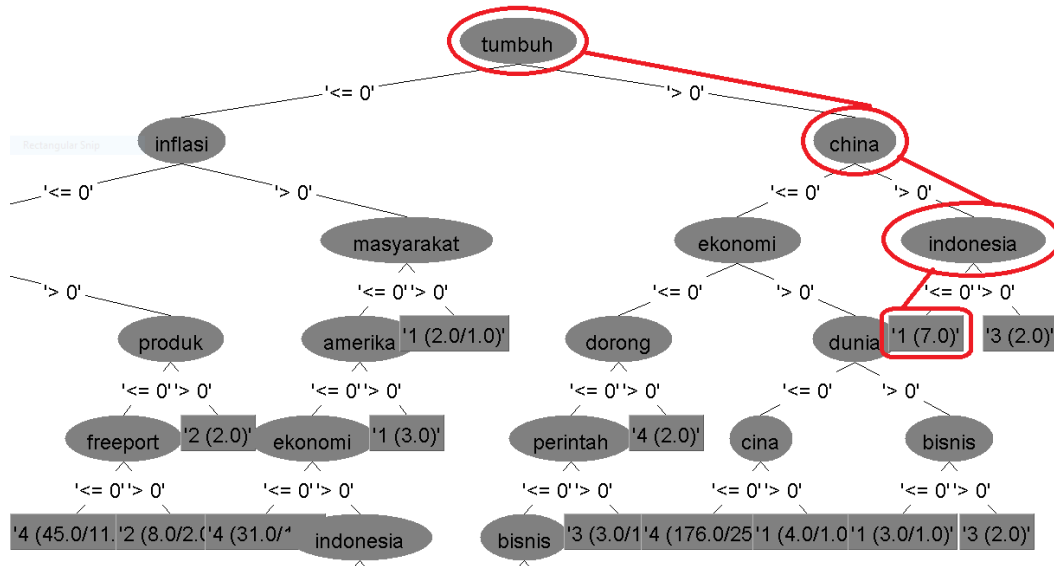
Status OK Log x0

Gambar 3-12 Hasil Pembelajaran Menggunakan *Decision Tree*



3. *Node* selanjutnya adalah “indonesia”. Pada contoh *tweet* tidak terdapat kata “indonesia” sehingga *tweet* tersebut diklasifikasikan sebagai kelas 1.

Ilustrasi penelusuran pada pohon keputusan seperti pada gambar 3-14.



Gambar 3-14 Klasifikasi Menggunakan Pohon Keputusan

Seluruh *tweet* pada *data training* dilakukan penelusuran menggunakan pohon keputusan tersebut. Proses utama dalam pembentukan pohon keputusan adalah menentukan atribut yang akan menjadi *node* terlebih dahulu. Proses tersebut menggunakan nilai *gain* pada tiap atribut. Proses penentuan *node* adalah dengan menentukan *gain* yang terbesar dari tiap atribut.

Sebagai contoh pembentukan pohon keputusan perlu dihitung *entropy* terlebih dahulu. Contoh *dataset* yang digunakan pada simulasi penghitungan ini seperti pada tabel 3.3.

Tabel 3.3 Contoh *Dataset Tweet* yang Telah Diklasifikasikan Secara Manual

No.	<i>Tweet</i>	Kelas
1.	persen tumbuh ekonomi rendah	3
2.	ekonomi bank tumbuh	3
3.	ekonomi masyarakat rendah bank tumbuh	3
4.	bank tumbuh persen potensi tumbuh	4
5.	potensi tumbuh ekonomi persen	4
6.	potensi ekonomi persen rendah	4
7.	masyarakat ekonomi bank	4

Langkah pertama adalah menjadikan *tweet* tersebut menjadi atribut yaitu kata. Hasil pemecahannya seperti pada tabel 3.4.

Tabel 3.4 Konversi *Tweet* ke Atribut

No.	Bank	Ekonomi	Masyarakat	persen	rendah	tumbuh	potensi	kelas
1.	0	1	0	1	1	1	0	3
2.	1	1	0	0	0	1	0	3
3.	1	1	1	0	1	1	0	3
4.	1	0	0	1	0	1	1	4
5.	0	1	0	1	0	1	1	4
6.	0	1	0	1	1	0	1	4
7.	1	1	1	0	0	0	0	4

Angka 0 dan 1 pada tabel 3.4 memiliki arti :

- 0 adalah tidak ada kata tersebut pada *tweet*
- 1 adalah terdapat kata tersebut pada *tweet*

Penentuan *node* pertama adalah dengan mencari nilai *gain* tertinggi pada 7 atribut yang tersedia. Proses pencarian *node* :

1. Dilakukan pencarian *entropy* total keseluruhan atribut.

Rumus *entropy* secara umum adalah :

$$Entropy (S) = \sum_{j=1}^k -p_j \log_2 p_j$$

Pada contoh kasus di tabel 3.4 terdapat 2 kelas yang dijadikan klasifikasi, sehingga nilai  $j = 1$  sampai 2. Sehingga persamaan dijabarkan menjadi :

$$\begin{aligned}
 Entropy(S) &= \sum_{j=1}^2 -p_j \log_2 p_j \\
 &= \left( -\left(\frac{3}{7}\right) \times \log_2 \left(\frac{3}{7}\right) \right) + \left( -\left(\frac{4}{7}\right) \times \log_2 \left(\frac{4}{7}\right) \right) \\
 &= 0,5239 + 0,4614 \\
 &= 0,9853
 \end{aligned}$$

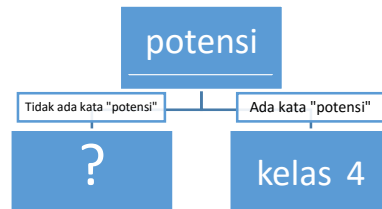
2. Dicari *entropy* pada setiap atribut dengan penghitungan yang sama dengan *entropy* total namun dibedakan setiap atribut dan kategorinya. Dari penghitungan tersebut diperoleh hasil seperti pada tabel 3.5.
- Tabel 3.5 Penghitungan *Gain* tiap Atribut

Atribut	Kategori	Jumlah	Jumlah kelas 3	Jumlah kelas 4	Entropy	Gain
Bank	1	4	2	2	1	0,0202
	0	3	1	2	0,9183	
Ekonomi	1	6	3	3	1	0,1281
	0	1	0	1	0	
Masyarakat	1	2	1	1	1	0,0060
	0	5	2	3	0,9710	
Persen	1	4	1	3	0,8113	0,1281
	0	3	2	1	0,9183	
Rendah	1	3	2	1	0,9183	0,1281
	0	4	1	3	0,8113	
Tumbuh	1	5	3	2	0,9710	0,2917
	0	2	0	2	0	
Potensi	1	3	0	3	0	<b>0,5216</b>
	0	4	3	1	0,8113	



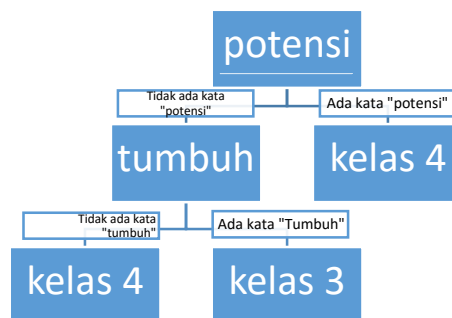
$$\begin{aligned}
 \text{Gain (Bank)} &= \text{Entropy (S)} - \sum_{i=1}^2 \frac{|S_i|}{|S|} \times \text{Entropy (S)} \\
 &= 0,9853 - \left( \left( \frac{|4|}{|7|} \times 1 \right) + \left( \frac{|3|}{|7|} \times 0,9183 \right) \right) \\
 &= 0,0202
 \end{aligned}$$

3. Nilai *gain* terbesar dimiliki oleh atribut “potensi” sehingga atribut tersebut menjadi *node* pertama dari pohon keputusan. Ilustrasi pohon keputusan yang terbentuk saat ini seperti pada gambar 3-15.



Gambar 3-15 *Node* dan *Leaf* Pertama yang Terbentuk

4. Setelah terbentuk *node* dan *leaf* pertama maka dibutuhkan *node* dan *leaf* selanjutnya. Proses penghitungan sama dengan menentukan *node* dan *leaf* pertama namun dengan menghilangkan atribut “potensi” dan data yang mengandung kata “potensi”. Hal ini dikarenakan atribut dan data tersebut sudah diklasifikasikan menjadi kelas 4 sesuai dengan pohon keputusan yang terbentuk. Bentuk akhir pohon keputusan yang dihasilkan seperti pada gambar 3-16.



Gambar 3-16 Pohon Keputusan yang Terbentuk

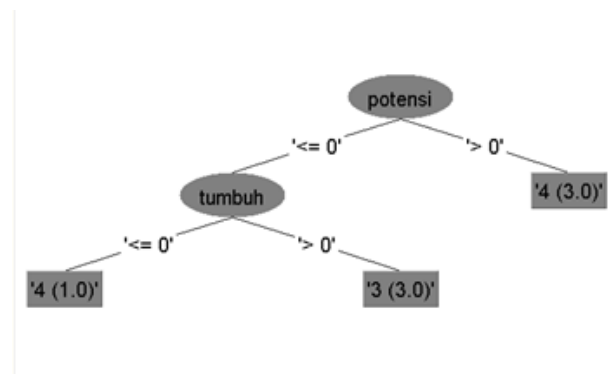
Data simulasi tersebut dicoba dimasukkan ke pohon keputusan yang sudah terbentuk. Salah satu contoh data *tweet* yaitu “ekonomi masyarakat rendah bank

tumbuh”. Pengecekan pertama adalah dengan melihat keberadaan kata “potensi” yang tidak terdapat pada *tweet* tersebut sehingga masuk ke *node* selanjutnya. Kata “tumbuh” terdapat pada *tweet* tersebut sehingga *tweet* tersebut dimasukkan ke kelas 3. Ilustrasi penentuan kelas pada *tweet* tersebut seperti pada gambar 3-17.



Gambar 3-17 Penentuan Kelas Menggunakan Pohon Keputusan

Dengan data simulasi yang sama dimasukkan ke aplikasi WEKA. Hasil pohon keputusan yang dibentuk oleh WEKA seperti pada gambar 3-18. Hasil yang terbentuk memiliki kesamaan dengan cara penghitungan manual.

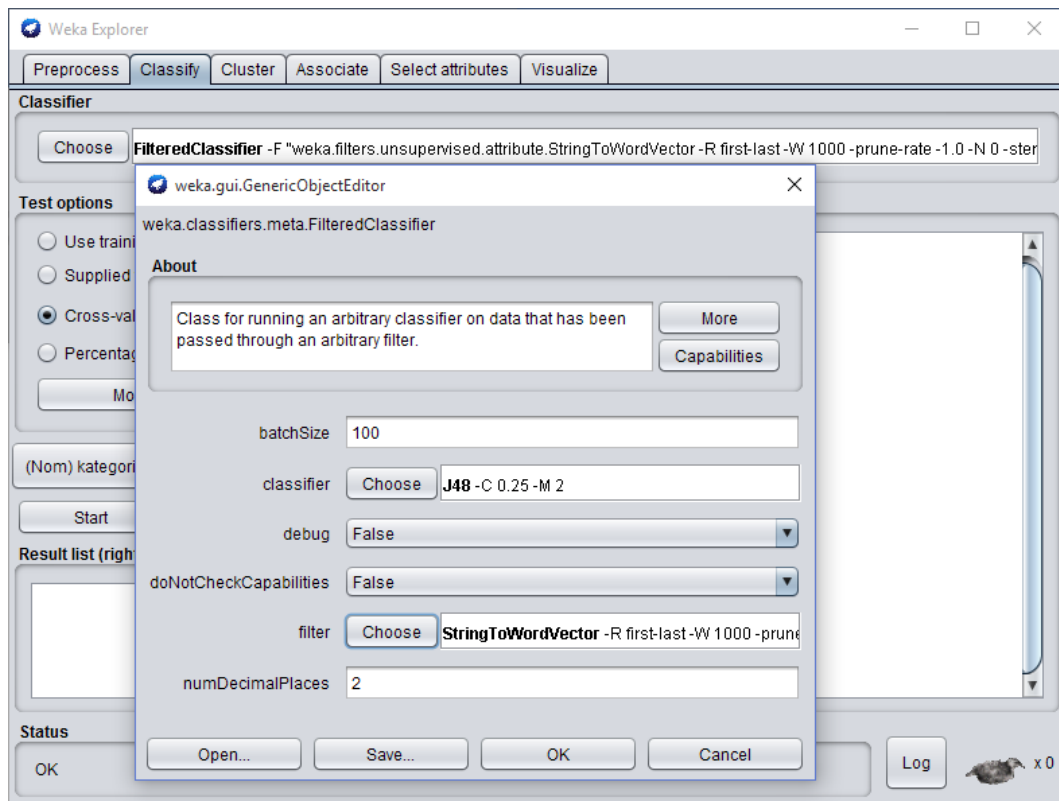


Gambar 3-18 Pohon Keputusan Hasil Proses di WEKA

### 3.9 Klasifikasi

Pohon keputusan yang telah terbentuk dari pembelajaran tersebut digunakan untuk melakukan klasifikasi terhadap keseluruhan *tweet*. Klasifikasi ini juga dilakukan pada WEKA menggunakan fasilitas *FilteredClassifier*. Penggunaan *FilteredClassifier* dikarenakan pada *classifier* tersebut memungkinkan untuk dilakukan klasifikasi terhadap data yang di filter terlebih dahulu. Pada proses ini *classifier Decision tree* digunakan untuk mengklasifikasikan data yang telah difilter

menggunakan `StringToWordVector`. Pilihan-pilihan yang terdapat pada *FilteredClassifier* terlihat pada gambar 3-19.



Gambar 3-19 Penggunaan *FilteredClassifier* pada WEKA

Data tes menggunakan file ARFF yang berisi keseluruhan *tweet* termasuk juga yang telah terpilih sebagai data training. Format pada *data training* dan *data test* harus sama, dalam hal ini adalah *tweet* dan kategori. Pada *data training* kategori sudah tercantum berdasarkan hasil klasifikasi manual. Sedangkan pada *data test* kategori diberi tanda “?” yang menandakan belum terklasifikasi. Gambar 3.20 menunjukkan contoh tampilan asli *data test*.

```

@relation 'Data Test 2'

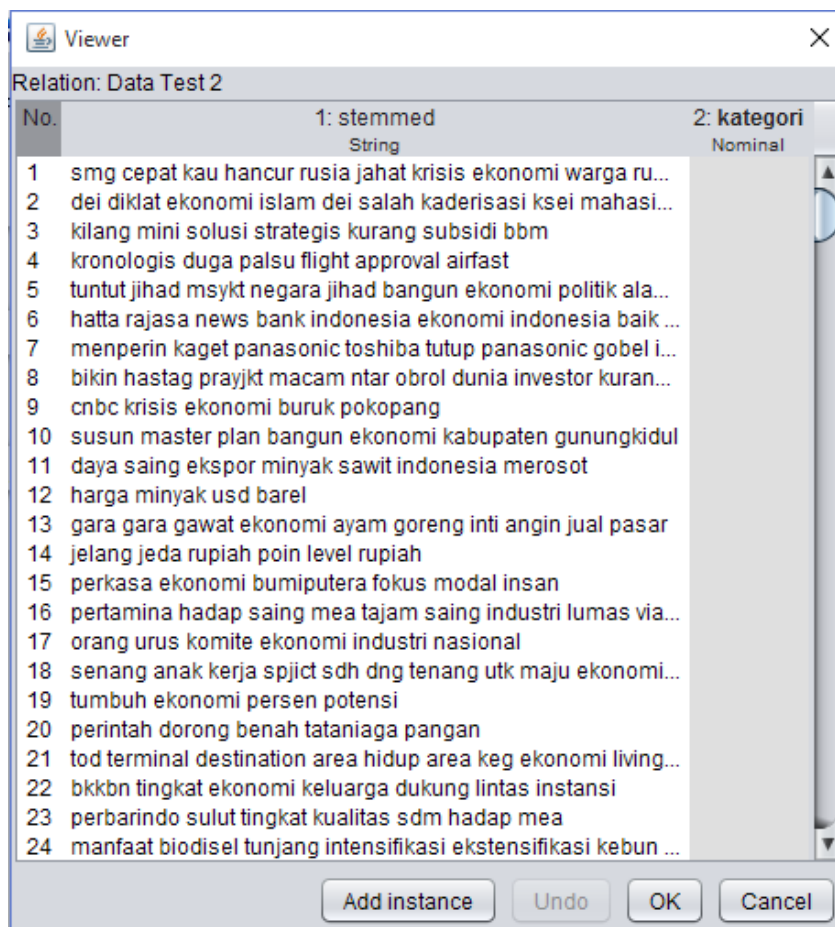
@attribute stemmed string
@attribute kategori {1,2,3,4}

@data
'smg cepat kau hancur rusia jahat krisis ekonomi warga rusia antri makan gratis',?
'dei diklat ekonomi islam dei salah kaderisasi ksei mahasiswa cont',?
'kilang mini solusi strategis kurang subsidi bbm',?
'kronologis duga palsu flight approval airfast',?
'tuntut jihad msykt negara jihad bangun ekonomi politik alam sihat adil bebas pimpin',?
'hatta rajasa news bank indonesia ekonomi indonesia baik tribunnewsbi ekonomi indones',?
'menperin kaget panasonic toshiba tutup panasonic gobel indonesia toshiba indonesia isu tel',?
'bikin hastag prayjkt macam ntar obrol dunia investor kurang rupiah ribu eko',?
'cnbc krisis ekonomi buruk pokopang',?
'susun master plan bangun ekonomi kabupaten gunungkidul',?
'daya saing ekspor minyak sawit indonesia merosot',?
'harga minyak usd barel',?
'gara gara gawat ekonomi ayam goreng inti angin jual pasar',?
'jelang jeda rupiah poin level rupiah',?
'perkasa ekonomi bumiputera fokus modal insan',?
'pertamina hadap saing mea tajam saing industri lumas via beritasatu',?
'orang urus komite ekonomi industri nasional',?
'senang anak kerja spjict sdh dng tenang utk maju ekonomi bangsa negara juang sejati',?

```

Gambar 3-20 Format *Data Test* yang Digunakan pada WEKA

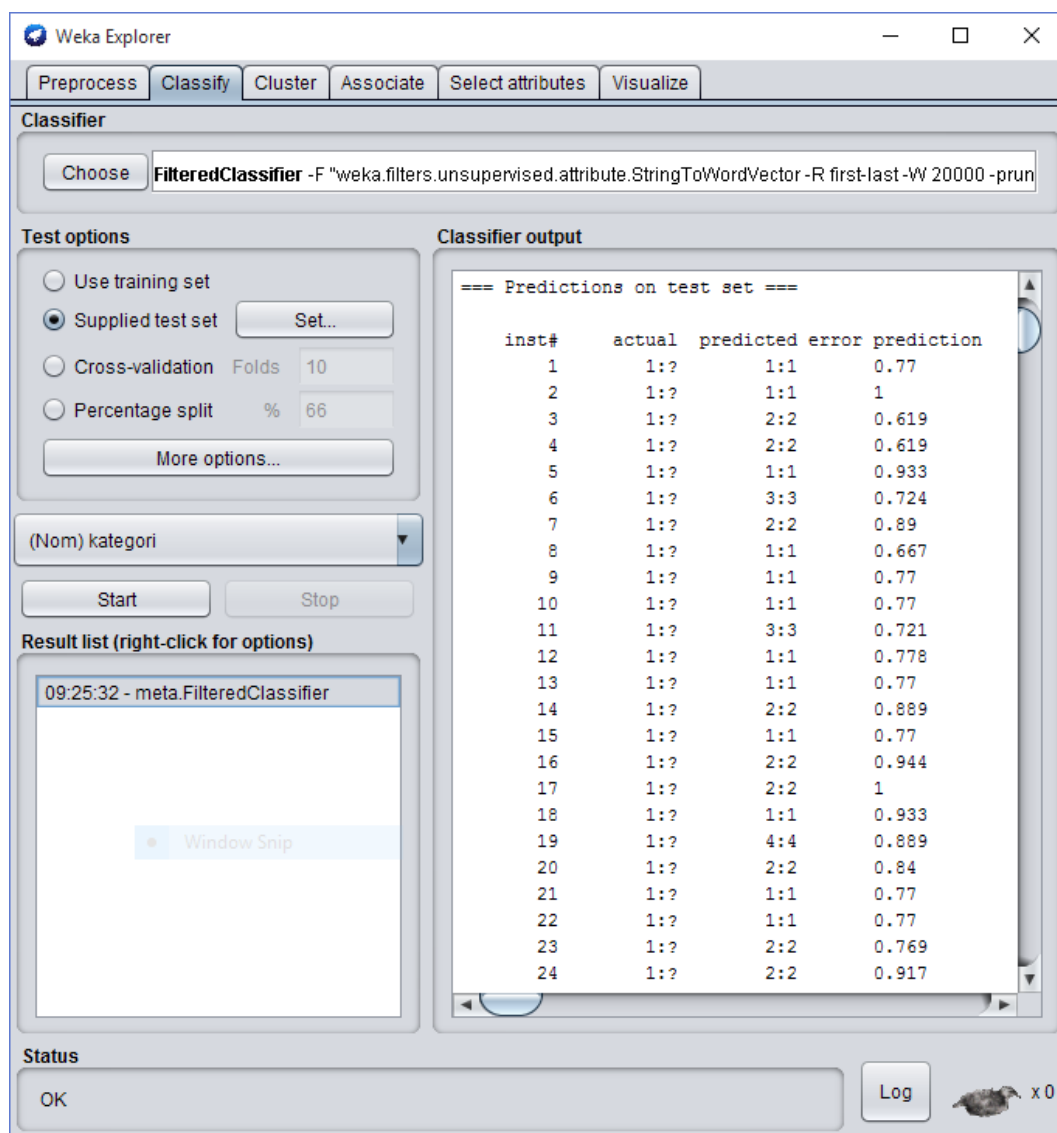
Gambar 3-21 menunjukkan tampilan *data test* pada WEKA. Pada gambar tersebut terlihat bahwa kolom kategori masih kosong. Hal tersebut menunjukkan bahwa *tweet* tersebut belum diklasifikasikan.



No.	1: stemmed String	2: kategori Nominal
1	smg cepat kau hancur rusia jahat krisis ekonomi warga ru...	
2	dei diklat ekonomi islam dei salah kaderisasi ksei mahasi...	
3	kilang mini solusi strategis kurang subsidi bbm	
4	kronologis duga palsu flight approval airfast	
5	tuntut jihad msykt negara jihad bangun ekonomi politik ala...	
6	hatta rajasa news bank indonesia ekonomi indonesia baik ...	
7	menperin kaget panasonic toshiba tutup panasonic gobel i...	
8	bikin hastag prayjkt macam ntar obrol dunia investor kuran...	
9	cnbc krisis ekonomi buruk pokopang	
10	susun master plan bangun ekonomi kabupaten gunungkidul	
11	daya saing ekspor minyak sawit indonesia merosot	
12	harga minyak usd barel	
13	gara gara gawat ekonomi ayam goreng inti angin jual pasar	
14	jelang jeda rupiah poin level rupiah	
15	perkasa ekonomi bumiputera fokus modal insan	
16	pertamina hadap saing mea tajam saing industri lumas via...	
17	orang urus komite ekonomi industri nasional	
18	senang anak kerja spjict sdh dng tenang utk maju ekonomi...	
19	tumbuh ekonomi persen potensi	
20	perintah dorong benah tataniaga pangan	
21	tod terminal destination area hidup area keg ekonomi living...	
22	bkkbn tingkat ekonomi keluarga dukung lintas instansi	
23	perbarindo sulut tingkat kualitas sdm hadap mea	
24	manfaat biodisel tunjang intensifikasi ekstensifikasi kebun ...	

Gambar 3-21 Format *Data Test* yang Terbaca oleh WEKA

Hasil proses ini merupakan prediksi klasifikasi *tweet* pada data tes. Seluruh *tweet* pada data tes diprediksi masuk ke salah satu kelas, sehingga keseluruhan *tweet* sudah memiliki kelas masing-masing. Gambar 3-22 menunjukkan contoh tampilan hasil klasifikasi terhadap data *test*.



Gambar 3-22 Contoh Hasil Klasifikasi pada Data *Test* Menggunakan WEKA

Hasil klasifikasi ditunjukkan pada gambar di atas. Kolom inst# adalah nomor *tweet*, sedangkan prediksi ditunjukkan pada kolom predicted. Hasil klasifikasi tersebut dipindahkan ke *worksheet* untuk mempermudah proses selanjutnya. Untuk mendapatkan bentuk sesuai dengan yang dibutuhkan, perlu

dilakukan modifikasi pada kolom di *worksheet*. Gambar 3-23 adalah contoh tampilan pada *worksheet* yang sudah disesikan dengan kebutuhan.

Tweet	inst#	actual	predicted	prediction
bank indonesia rate turun ekonomi indonesia proyeksi tumbuh bank indonesia rate turun giro	52429	1:?	4:4	0,889
cemas perban turun suku bunga jusef kalla ekonomi daya saing	52430	1:?	2:2	0,75
maskapai rebut kue wisatawan muslim	52431	1:?	2:2	0,619
tingkat kredit konsumen bank mandiri incar kta kpr	52432	1:?	2:2	0,769
desak turun suku bunga kelola uang negara revisi	52433	1:?	1:1	0,545
tumbuh ekonomi china rendah	52434	1:?	1:1	1
jokowi gandeng facebook wujud ekonomi digital indonesia	52435	1:?	2:2	0,683
menkeu masyarakat bantu bangun obligasi ritel	52436	1:?	2:2	0,619
dampak ekonomi lemah mea siaga	52437	1:?	3:3	0,757
areal kelapa sawit indonesia juta produksi cpo juta ton nilai ekspor juta	52438	1:?	3:3	0,721
harga beras daging cabe dll Selasa Januari	52439	1:?	2:2	0,879
gratis distribusi pupuk cair enero penuh target beritajatim	52440	1:?	2:2	0,619
startup ukm bantu tumbuh ekonomi	52441	1:?	4:4	0,889
cikgu pandai pakat doktor tinggal bijak ambil ekonomi menteri	52442	1:?	1:1	0,77
usaha Eropa optimistis ekonomi Indonesia	52443	1:?	2:2	0,683

Gambar 3-23 Format *Worksheet* Hasil Konversi dari Klasifikasi

### 3.10 Identifikasi Topik Utama

Hasil proses klasifikasi berupa *tweet* yang telah dibagi tiap kelas tersebut kemudian dianalisis. Tahap pertama analisis adalah melakukan identifikasi topik utama pada setiap kelas. Proses selanjutnya adalah mencari topik utama pada tiap kelas *tweet* yang membahas perekonomian. Untuk mendapatkan topik yang tepat maka digunakan n-gram. Jenis n-gram yang digunakan adalah unigram, bigram, dan trigram. N-gram pada tiap kelas akan diurutkan berdasarkan frekuensi terbanyak kemunculannya. Hal ini dilakukan untuk mengetahui n-gram yang paling sering digunakan pada tiap kelas. Selanjutnya n-gram tersebut akan digunakan sebagai kandidat topik utama.

Pencarian n-gram ini memanfaatkan Microsoft Excel dan kode VBA. Gambar 3.24 menunjukkan contoh hasil pencarian n-gram pada *worksheet*.

Tweet	Unigram	#	Bigram	#	Trigram	#
sng cepat kau hancur rusia jahat krisis ekonomi warg ekonomi	tumbuh	1379	tumbuh ekonomi	212	ekonomi indonesia baik	30
dei diklat ekonomi islam dei salah kaderisasi ksei ma indonesia	tumbuh	468	ekonomi indonesia	128	ekspor minyak sawit	28
kronologis duga palsu flight approval airfast	tumbuh	302	bank indonesia	115	bank indonesia ekonomi	22
tuntut jihad msykt negara jihad bangun ekonomi poli ekspor	ekonomi	187	ekonomi lemah	87	bank indonesia rate	22
bikin hastag prayjkt macam ntar obrol dunia investor bank	bank	172	harga minyak	50	indonesia ekonomi indonesia	22
gara gara gawat ekonomi ayam goreng inti angin jual harga	harga	138	fakultas ekonomi	44	tumbuh ekonomi indonesia	21
perkasa ekonomi bumiputera fokus modal insan lemah	lemah	121	dolar amerika	41	inflasi alami baik	20
culi bunuh anak depok duga motif ekonomi sawit	sawit	119	minyak sawit	39	paket bijak ekonomi	20
tumbuh ekonomi cina lamban	inflasi	117	data ekonomi	36	baik inflasi alami	18
follow sumber berita basi basi olahraga sehat ekonomi perintah	perintah	115	kelapa sawit	35	indonesia baik inflasi	18
kuliah hujan fakultas ekonomi unnm pic	turun	113	kereta cepat	35	dolar amerika merosot	16
perintah negara mana dunia hantam uang ekonomi c minyak	minyak	110	indonesia baik	30	masyarakat ekonomi asean	16
dar krisis ekonomi payah laa ikan bilis ekor	rupiah	109	jusuf kalla	30	data ekonomi lemah	15
data tumbuh ekonomi china pasar rilis data tumbuh usaha	usaha	103	ekspor minyak	29	indonesia bank indonesia	15
ekonomi perintah rencana	baik	88	kondisi ekonomi	27	agus marto ekonomi	14
polres pamekasan bentuk tim khusus narkoba warta jokowi	jokowi	85	warta ekonomi	27	proyek kereta cepat	14
ktp slip gaji aju kredit card bank mega jasa	tingkat	84	ekonomi china	26	turun suku bunga	14
dilm agama ade jenis najis najis mukhafafah najis mul industri	industri	77	krisis ekonomi	26	andal ekspor barang	13
tahniah politik moga allah kati juang martabat ekono negara	negara	72	sawit indonesia	26	bank indonesia bank	13
imf pangkas proyeksi tumbuh ekonomi global	amerika	70	bijak ekonomi	25	bank indonesia prediksi	13
kelas kah pasar fakultas ekonomi bisnis unlam	menteri	68	ekonomi tumbuh	24	daya saing ekspor	13
cuit bareskrim panjang tahan ongen direktorat tindak pasar	pasar	67	indonesia ekonomi	23	dorong tumbuh indonesia	13

Find Density

Clear

Gambar 3-24 Tampilan Identifikasi Unigram, Bigram dan Trigram Berdasarkan Frekuensi Kemunculan

Contoh kode VBA yang digunakan untuk mencari n-gram seperti pada gambar 3-25.

```

Sub PhraseDensity(nWds As Long, Col As Variant)
    Dim astr() As String
    Dim i As Long
    Dim j As Long
    Dim cell As Range
    Dim sPair As String
    Dim rOut As Range

    With CreateObject("Scripting.Dictionary")
        .CompareMode = vbTextCompare
        For Each cell In Range("A1", Cells(Rows.Count, "A").End(xlUp))
            astr = Split(Letters(cell.Value), " ")

            For i = 0 To UBound(astr) - nWds + 1
                sPair = vbNullString
                For j = i To i + nWds - 1
                    sPair = sPair & astr(j) & " "
                Next j
                sPair = Left(sPair, Len(sPair) - 1)

                If Not .exists(sPair) Then
                    .Add sPair, 1
                Else
                    .Item(sPair) = .Item(sPair) + 1
                End If
            Next i
        Next cell

        Set rOut = Columns(Col).Resize(.Count, 2).Cells
        rOut.EntireColumn.ClearContents

        rOut.Columns(1).Value = Application.Transpose(.Keys)
        rOut.Columns(2).Value = Application.Transpose(.Items)

        rOut.Sort Key1:=rOut(1, 2), Order1:=xlDescending, _

```

Gambar 3-25 Kode VBA untuk Identifikasi Unigram, Bigram dan Trigram

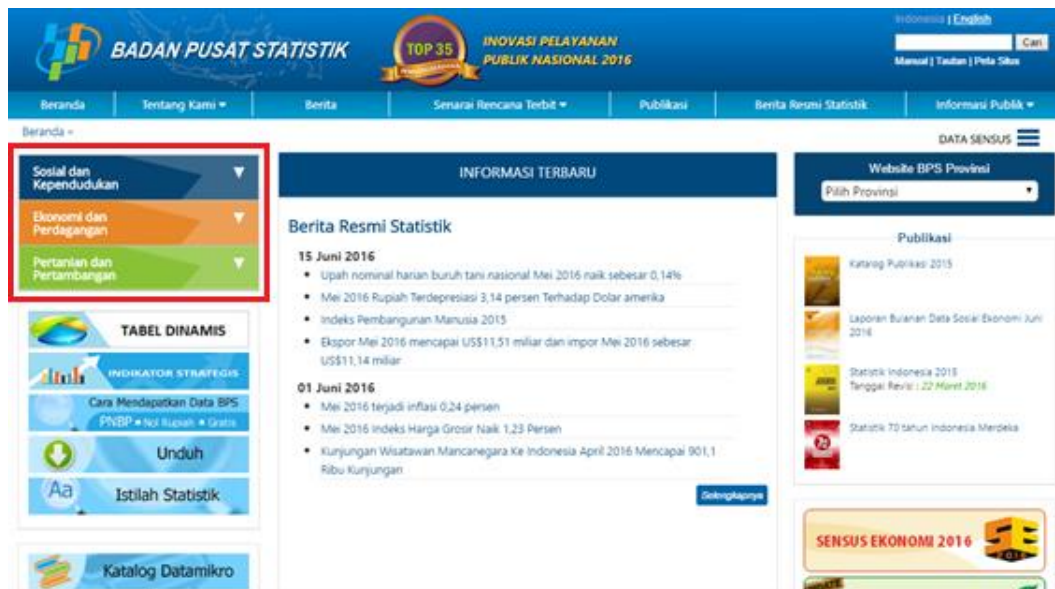
Hasil pencarian n-gram dengan frekuensi kemunculan paling banyak tersebut digunakan sebagai kandidat topik utama pada setiap kelas. Kandidat topik tersebut yang akan menjadi bahan analisis pada penelitian ini.

### 3.11 Data BPS Bidang Perekonomian

Analisa pada penelitian ini menghubungkan antara topik utama yang paling sering muncul pada *tweet* dengan ketersediaan data BPS. Oleh karena itu diperlukan daftar data-data pada bidang perekonomian yang disediakan oleh BPS. Daftar data yang tersedia tersebut tercantum pada halaman utama website resmi BPS. Data-data yang disediakan dibagi menjadi 3 kelompok utama yaitu :

1. Sosial dan Kependudukan
2. Ekonomi dan perdagangan
3. Pertanian dan Pertambangan

Data pada kelompok ekonomi dan perdagangan adalah yang paling sesuai dengan topik pembahasan pada penelitian ini. Sehingga data pada topik tersebut akan digunakan pada pembahasan di penelitian ini. Gambar 3.26 menunjukkan tampilan utama website BPS yang menampilkan kelompok data yang dihasilkan.



Gambar 3-26 Tampilan Halaman Utama Website Badan Pusat Statistik



## BAB 4

### HASIL DAN PEMBAHASAN

#### 4.1 Pengumpulan *Tweet*

Pengumpulan *tweet* dilakukan pada waktu 2 bulan yaitu pada bulan januari dan februari 2016. Hasil yang didapatkan adalah sebanyak 33.497 *tweet* yang mengandung kata kunci awal yaitu “ekonomi”. *Tweet* yang dikumpulkan tersebut sudah dihilangkan dari duplikasi. Hal ini dilakukan untuk membersihkan data dari *retweet* yang bisa memungkinkan hasil yang menjadi bias. Pengumpulan *tweet* menggunakan alat Tweet Archiver tersebut menghasilkan 16 atribut untuk tiap *tweet*. Atribut yang didapatkan tersebut pada tabel 4.1.

Tabel 4.1 Atribut yang Berhasil Dikumpulkan oleh Twitter Archiver

Nama atribut	Nama atribut	Nama atribut	Nama atribut
Date	<i>Tweet ID</i>	<i>Retweets</i>	Location
Screen Name	App	Favorites	Bio
Full Name	Followers	Verified	Profile Image
<i>Tweet Text</i>	Follows	User Since	Google Maps

Dari atribut yang berhasil dikumpulkan tersebut hanya 1 atribut yang digunakan pada penelitian ini yaitu *Tweet Text*. Atribut tersebut adalah isi dari *tweet* yang sesuai dengan filter kata kunci awal yang digunakan. Kata kunci awal yang digunakan cukup umum, oleh karena tidak semua *tweet* hasil pengumpulan sesuai dengan harapan. Sebagai contoh adalah *tweet* yang menggunakan bahasa malaysia yang mirip dengan bahasa indonesia. Terdapat juga *tweet* yang hanya mengandung *Check-in* pada suatu tempat. Beberapa *tweet* yang terkumpul juga mengandung karakter-karakter yang tidak dibutuhkan. Sebagian besar *tweet* juga mengandung *hyperlink*.

#### 4.2 Klasifikasi Manual

Proses klasifikasi manual ini dilakukan untuk membentuk data latih. Pada penelitian ini digunakan 1.858 *tweet* sebagai data latih. Pemilihan *tweet* untuk data

latih ini berdasarkan pilihan acak karena pada saat ini belum terdapat informasi mengenai pengelompokan *tweet*. Pada tahap ini diketahui sebagian besar *tweet* menyertakan *hyperlink* pada isi *tweet*nya. Hal tersebut mengindikasikan bahwa sebagian besar *tweet* yang terkumpul memiliki acuan yang jelas yaitu *hyperlink* yang disertakan. Jadi sebagian besar *tweet* bukan hanya sekedar kicauan pengguna namun lebih acuan yaitu *hyperlink* tersebut. Pada kenyataannya seringkali penentuan kelas terhadap data latih membutuhkan pendalaman pada *hyperlink* yang disertakan. Hasil klasifikasi manual seperti tersebut pada tabel 4.2.

Tabel 4.2 Hasil Klasifikasi *Tweet* secara Manual

<b>Nama Kelas</b>	<b>Deskripsi Kelas</b>	<b>Jumlah</b>
Kelas Pertama	<i>Tweet</i> yang tidak termasuk topik perekonomian negara indonesia	681
Kelas Kedua	<i>Tweet</i> yang memiliki topik perekonomian negara namun tidak menyinggung/menggunakan data BPS	753
Kelas Ketiga	<i>Tweet</i> yang memiliki topik perekonomian negara dan menyinggung/menggunakan data yang mungkin bersumber dari BPS	169
Kelas Keempat	<i>Tweet</i> yang memiliki topik perekonomian negara dan menyinggung/menggunakan data yang bersumber dari BPS	255
<b>Total</b>		<b>1.858</b>

Dari tabel 4.2 terlihat bahwa kelas yang dibentuk ada 4 kelas. Kelas pertama merupakan kumpulan *tweet* yang diluar topik pembahasan penelitian ini. Kelas ini dibentuk karena diperkirakan terdapat *tweet* yang tidak sesuai dengan penelitian ini.

### 4.3 Pra Proses

Setelah dilakukan proses klasifikasi manual maka proses selanjutnya adalah melakukan pra proses. Pada tahap ini terdapat beberapa proses yang dilakukan yaitu :

1. *Cleaning*
2. *Case folding*
3. *Stopword removal*
4. *Stemming*

Contoh hasil pra proses pada *tweet* ditampilkan pada tabel 4.3.

Tabel 4.3 Contoh *Tweet* yang telah Dilakukan tahap Pra Proses

Teks <i>Tweet</i>	Kelas
ekonomi lambat jufuf kalla temu gantung asing	4
sukfes trs oki pulp paper sdh mendpat jalur hijau kontribusi thdp tingkat tumbuh ekonomi	4
relaksasi ekspor mineral mentah pihak freeport	3
usaha sawit undang undang revisi perintah anggap tuju bakar hutan indonesia	2
ayo kembang umkm sbgai gerak ekonomi lbh	3
pede ekonomi tumbuh mdlm pasang target via kontannews	4
bank indonesia laju ekonomi pulau jawa salip nasional	3
kuliah hujan fakultas ekonomi unnm pic	1
data tumbuh ekonomi china pasar rilis data tumbuh pdb china kuartal	1
pakde karwo tawar jatimnomics tumbuh ekonomi nasional	4
newsvideo ekonom optimistis tumbuh ekonomi baik	4
menkeu optimis indonesia masuk negara ekonomi besar dunia	3

### 4.4 Pembelajaran

Tahap selanjutnya adalah proses pembelajaran menggunakan data latih yang telah diklasifikasikan secara manual. Proses pembelajaran data latih dilakukan pada *software* WEKA. Metode yang diuji coba pada data latih adalah Naive Bayes

dan *Decision tree*, yang mana metode Naive Bayes dianggap sebagai dasar perbandingan akurasi model. Beberapa kombinasi yang digunakan pada percobaan Naive Bayes adalah *minimum terms frequency* atau bisa disebut minimal cacah kata. Batasan ini memiliki makna bahwa jumlah minimal sebuah kata yang muncul pada himpunan kata. Hasil percobaan menggunakan metode Naive Bayes tersebut pada tabel 4.4.

Tabel 4.4 Tingkat Akurasi Model yang Dibentuk Menggunakan Metode Naive Bayes

<b>Jumlah Minimal Cacah Kata (<i>Minimum Term Frequency</i>)</b>	<b>Nilai Akurasi</b>
1	65,71
2	65,66
3	65,60
4	65,50
5	66,09
6	65,50
7	65,66
8	65,33
9	65,33
10	65,17

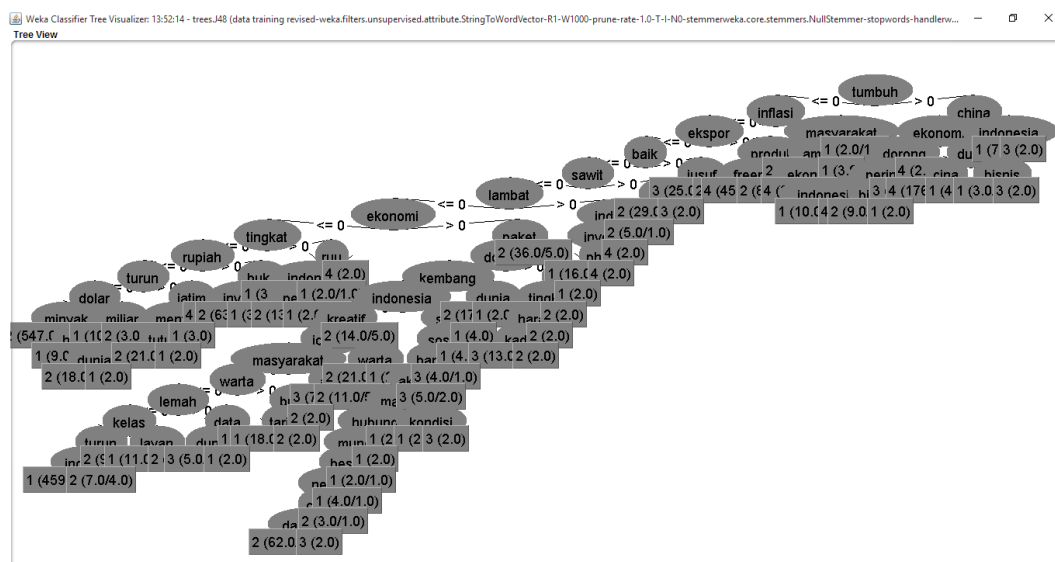
Pada penggunaan metode Naive Bayes nilai akurasi tertinggi diperoleh dengan menggunakan minimal cacah kata 5 dengan akurasi 66,09 %. Hal tersebut berarti akurasi optimal dengan metode Naive Bayes diperoleh dengan menghilangkan kata-kata yang muncul dibawah 5 kali pada keseluruhan data.

Sedangkan hasil percobaan menggunakan metode *Decision Tree* ditunjukkan pada tabel 4.5.

Tabel 4.5 Tingkat Akurasi Model yang dibentuk Menggunakan Metode *Decision Tree*

Confidence Factor	Min Term Freq 3	Min Term Freq 4	Min Term Freq 5	Min Term Freq 6	Min Term Freq 7
0,2	67,76	67,38	67,97	67,43	67,54
0,25	67,54	67,27	67,59	67,43	67,33
0,3	67,27	67,22	<b>68,03</b>	67,38	67,59
0,35	66,95	67,65	67,33	67,27	67,43
0,4	66,73	66,46	67,11	66,52	66,89

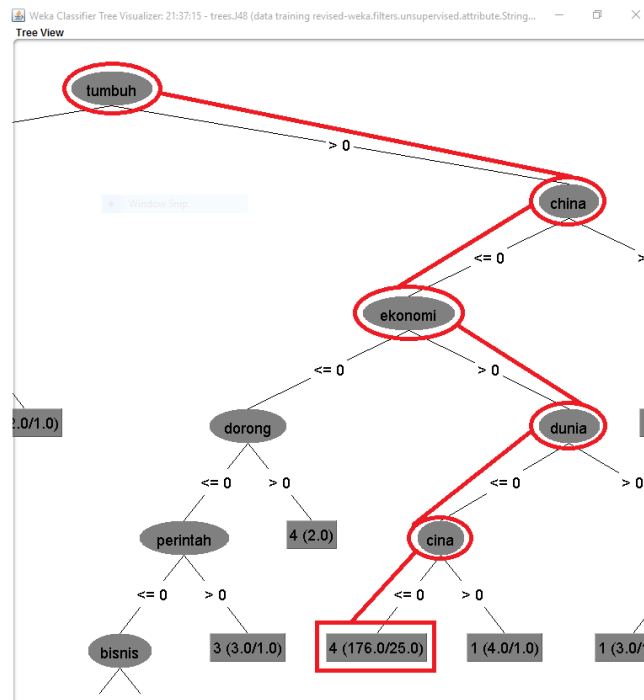
Berdasarkan tabel 4.4 dan tabel 4.5 terlihat bahwa penggunaan *minimum terms frequency* yang bervariasi menghasilkan tingkat akurasi model yang berbeda-beda. Berdasarkan percobaan dengan 2 kombinasi metode dan pilihan maka didapatkan akurasi optimum sebesar 68,03 % dengan menggunakan metode *Decision tree* dengan *minimum terms frequency* 5 dan *confidence factor* 0,3. Sedangkan pohon keputusan yang berhasil dibentuk memiliki 158 *node* dan 80 *leaf* seperti pada gambar 4.1.



Gambar 4-1 Pohon Keputusan yang Berhasil Dibentuk

Salah satu contoh *tweet* yang diklasifikasikan adalah “rendah tumbuh ekonomi indonesia triwulan banding”. Ketika *tweet* tersebut dimasukkan pada

pohon keputusan yang ada maka proses klasifikasinya melalui pohon keputusan seperti pada gambar 4.2.



Gambar 4-2 Proses Klasifikasi *Tweet* pada Pohon Keputusan yang Terbentuk

Proses pengklasifikasiian menggunakan pohon keputusan pada gambar 4-2 bercerita bahwa proses dimulai dengan pencarian *node* pertama dan dilanjutkan *node* berikutnya sampai dengan bertemu dengan *leaf*. Untuk *tweet* yang dijadikan contoh terebut pengklasifikasian dimulai dengan melihat keberadaan kata “tumbuh”, “china”, “ekonomi”, “dunia”, dan “china”. *Leaf* terakhir adalah kelas 4, hal itu menunjukkan bahwa tweet tersebut diklasifikasikan ke kelas 4.

Kemudian dilakukan pengecekan ke *worksheet* hasil klasifikasi. Pada *worksheet* hasil klasifikasi terlihat bahwa *tweet* tersebut juga diklasifikasikan pada kelas 4. Pengecekan hasil klasifikasi tersebut terlihat pada gambar 4-3.

	A	E
1	tweet	predicted
14	gara gara gawat ekonomi ayam goreng inti angin jual pasar	1:1
15	jelang jeda rupiah poin level rupiah	2:2
16	perkasa ekonomi bumiputera fokus modal insan	1:1
17	pertamina hadap saing mea tajam saing industri lumas via beritasatu	2:2
18	tumbuh ekonomi persen potensi	4:4
19	manfaat biodisel tunjang intensifikasi ekstensifikasi kebun sawit jamin pasok cpo	2:2
20	jusuf kalla gantung ekonomi indonesia orang okezone	2:2
21	pikir pikir senjang sosial ekonomi	1:1
22	rendah tumbuh ekonomi indonesia triwulan banding	4:4
23	culi bunuh anak depok duga motif ekonomi	1:1
24	bank indonesia ekonomi indonesia baik inflasi alami baik	4:4
25	mahasiswa kes ekonomi skrng	1:1
26	tumbuh ekonomi cina lamban	1:1

Gambar 4-3 Hasil Klasifikasi *Tweet* Contoh

#### 4.5 Klasifikasi

Model yang memiliki akurasi paling optimum digunakan untuk melakukan klasifikasi terhadap sisa *tweet* yang ada. Model yang digunakan adalah dengan menggunakan metode *Decision tree*, *minimum terms frequency* 5 dan *confidence factor* 0,3. Hasil klasifikasi menggunakan aplikasi WEKA seperti pada tabel 4.6.

Tabel 4.6 Ringkasan Hasil Prediksi *Tweet* Menggunakan Metode *Decision Tree*

Nama Kelas	Deskripsi	Jumlah <i>Tweet</i>	Persentase	Persentase terhadap topik perekonomian
Pertama	<i>Tweet</i> yang tidak termasuk topik perekonomian negara indonesia	14.252	42,55	
Kedua	<i>Tweet</i> yang memiliki topik perekonomian negara namun tidak menyinggung/menggunakan data BPS	16.479	49,20	85,63

Tabel 4.7 Ringkasan Hasil Prediksi *Tweet* Menggunakan Metode *Decision Tree* (Lanjutan).

Ketiga	<i>Tweet</i> yang memiliki topik perekonomian negara dan menyinggung/menggunakan data yang mungkin bersumber dari BPS	1.115	3,22	5,79
Keempat	<i>Tweet</i> yang memiliki topik perekonomian negara dan menyinggung/menggunakan data yang bersumber dari BPS	1.651	4,93	8,58
<b>Total</b>		<b>33.497</b>	<b>100,00</b>	<b>100,00</b>

Beberapa hal yang bisa dilihat dari hasil klasifikasi tersebut diantaranya :

1. Kelas pertama memiliki persentase yang cukup besar yaitu 42,55 % dari total *tweet*. Hal ini cukup membuktikan dugaan awal bahwa kelas ini perlu dibentuk. Kelas ini berisi *tweet* yang diluar topik perekonomian Indonesia. Contoh *tweet* yang diklasifikasikan pada kelas ini diantaranya :
  - a. tumbuh ekonomi cina lamban
  - b. turun jual ritel tegas lambat ekonomi amerika
  - c. tekan ekonomi
  - d. alhamdulillah company pgil kerja hihi grab mna ekonomi skrg xleh pilih heeee
  - e. selamat pagi fighter fighter perang ekonomi moga urus mudah
  - f. jurni emang anak ekonomi
2. Kelas kedua merupakan kelas yang mempunyai persentase *tweet* paling besar yaitu 49,20 % dari seluruh *tweet*. Kelas ini berisi *tweet* yang merupakan topik perekonomian namun tidak menggunakan



dan/atau berhubungan dengan data BPS. Jika *tweet* yang memiliki topik diluar perekonomian dikeluarkan maka kelas ini memiliki persentase 85,63 % dari *tweet* topik perekonomian. Beberapa contoh *tweet* yang diklasifikasikan pada kelas ini diantaranya :

- a. obat paket ekonomi jokowi mujarab
- b. agen asing lomba selundup poin liberalisasi paket ekonomi menteri keblinger
- c. usaha tambang saran core core bisnis emiten sektor tambang alami tekan
- d. intensifikasi kebun sawit topang sedia dana riset kembang produktivitas kebun
- e. panggil dirjen minerba freeport sepakat rdp komisi vii via metrotv

3. Kelas ketiga memiliki persentase 3,22 % dari total *tweet*. Jika *tweet* yang memiliki topik diluar perekonomian dikeluarkan maka kelas ini memiliki persentase 5,79 % dari *tweet* yang memiliki topik perekonomian. Beberapa contoh *tweet* yang diklasifikasikan pada kelas ini diantaranya :

- a. daya beli masyarakat turun laku ekonomi tanggung akibat dampak
- b. harap ekonomi indonesia baik monyet api
- c. menteri rizal ramli blok masela dorong ekonomi indonesia timur
- d. bank indonesia laju ekonomi pulau jawa salip nasional

4. Kelas keempat memiliki persentase 4,93 % dari total *tweet*. Jika *tweet* yang memiliki topik diluar perekonomian dikeluarkan maka kelas ini memiliki persentase 8,58 % dari *tweet* topik perekonomian. Beberapa contoh *tweet* yang diklasifikasikan pada kelas ini diantaranya :

- a. agus marto ekonomi tumbuh inflasi via
- b. dana desa rupiah sumbang tumbuh ekonomi nasional
- c. silah cek data bps sektor ikan mndrong tumbuh ekonomi bngsa hncurnya hrg komoditas

- d. bank indonesia optimis ekonomi indonesia tumbuh proyeksi ekonomi bank indonesia arah
- 5. Tingkat penggunaan data BPS pada *tweet* bidang perekonomian Indonesia diperkirakan sebesar 8,58 %. Angka ini belum bisa diartikan apa-apa karena saat ini belum angka pembandingan lain. Bahkan angka ini bisa dijadikan patokan awal untuk evaluasi BPS di masa depan
- 6. Kelas ketiga berisi *tweet* yang bisa menggunakan dan/atau berhubungan dengan data BPS. Sebagai contoh adalah ketika ada *tweet* yang berisi tentang daya beli, hal itu bisa saja berhubungan dengan data daya beli yang dimiliki oleh BPS. Namun tidak menutup kemungkinan bahwa *tweet* tersebut sama sekali tidak berhubungan dengan data daya beli yang dimiliki oleh BPS. Sebagai contoh lain ketika ada *tweet* yang menyebutkan perekonomian membaik, hal tersebut bisa berhubungan dengan data pertumbuhan ekonomi atau inflasi BPS. Namun bisa jadi yang dimaksud adalah membaiknya perekonomian dari sudut pandang penulis *tweet*.

#### **4.6 Identifikasi Kandidat Topik Utama**

Hasil klasifikasi pada kelas kedua akan diteliti lebih dalam lagi. Pada kelas pertama dan kelas ketiga tidak dilakukan analisis lebih lanjut karena kelas tersebut tidak memiliki sesuatu yang khusus. Pada kelas kedua dan keempat dilakukan identifikasi topik utama untuk mengetahui kemungkinan isu yang bisa digunakan sebagai masukan kepada BPS untuk penyempurnaan diseminasi data.

Tahap ini memanfaatkan penggunaan n-gram. N-gram yang digunakan pada analisis adalah bigram dan trigram. Unigram tidak digunakan untuk analisis lebih lanjut karena unigram kurang spesifik mengacu pada sebuah topik. Tabel 4.7 menunjukkan hasil identifikasi unigram pada kelas kedua berdasarkan frekuensi kemunculannya.

Tabel 4.8 Daftar Unigram berdasarkan Frekuensi Kemunculan

Unigram	Frekuensi		Unigram	Frekuensi
ekonomi	3.380		usaha	757
indonesia	2.922		jokowi	748
rupiah	1.500		industri	691
harga	1.359		via	672
perintah	1.123		paket	621
turun	911		cepat	574
bank	849		menteri	572

Unigram yang diurutkan berdasarkan frekuensi kemunculan tersebut belum memiliki arti yang spesifik pada suatu topik. Sebagai contoh ekonomi, indonesia, rupiah, perintah, harga, dan seterusnya merupakan kata yang sangat umum sehingga sangat sulit mencari topik yang sesuai.

Untuk mendapatkan topik yang lebih spesifik maka digunakan bigram dan trigram. Bigram pada kelas kedua berdasarkan frekuensi kemunculannya seperti tercantum pada tabel 4.8.

Tabel 4.9 Daftar Bigram berdasarkan Frekuensi Kemunculan

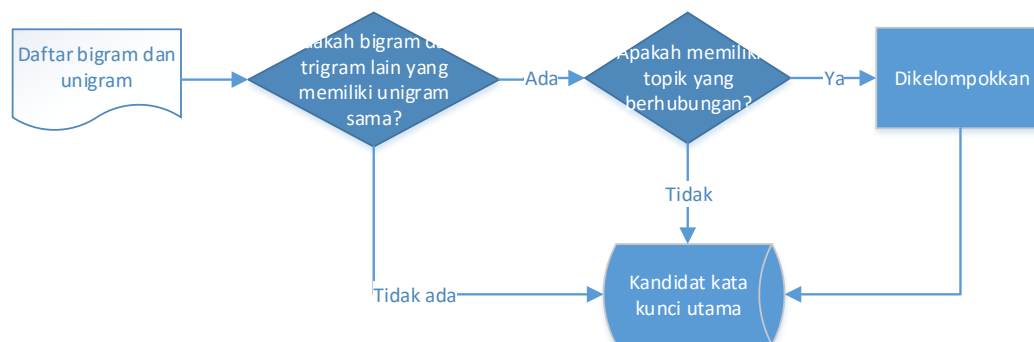
Bigram	Frekuensi		Bigram	Frekuensi
sawit indonesia	517		rupiah triliun	300
kereta cepat	392		masyarakat ekonomi	262
paket bijak	361		ekonomi asean	256
bijak ekonomi	352		republika online	226
bank indonesia	339		paket ekonomi	210
harga minyak	336		jusuf kalla	193
ekonomi kreatif	300		metro news	172

Sedangkan trigram yang berhasil diidentifikasi berdasarkan frekuensi kemunculan seperti pada tabel 4.9.

Tabel 4.10 Daftar Trigram berdasarkan Frekuensi Kemunculan

Trigram	Frekuensi	Trigram	Frekuensi
paket bijak ekonomi	322	bijak ekonomi jilid	77
masyarakat ekonomi asean	251	kereta cepat jakarta	77
proyek kereta cepat	156	nilai tukar rupiah	58
bank indonesia rate	102	turun suku bunga	56
ekonomi asean mea	91	indeks harga saham	55
cepat jakarta bandung	85	harga saham gabung	54
badan ekonomi kreatif	80	harga daging sapi	51

Tabel 4.8 dan tabel 4.9 menunjukkan masing-masing 14 bigram dan trigram yang paling sering muncul pada kelas kedua. Dari total 28 bigram dan trigram ini akan dilakukan identifikasi bigram dan trigram yang memiliki makna sama. Pencarian makna yang sama ini digunakan untuk meminimalisir kandidat topik utama. Diagram alir yang digunakan untuk identifikasi kesamaan topik bigram dan trigram seperti pada gambar 4.4.



Gambar 4-4 Alur Kerja Pengelompokan N-gram

Tabel 4.11 Hubungan antar N-gram dan Penentuan Kandidat Topik Utama

<b>N-gram</b>	<b>Hubungan dengan n-gram lain</b>	<b>Nama Kandidat Topik Utama Terpilih</b>
sawit indonesia	-	Sawit indonesia
kereta cepat	Proyek kereta cepat kereta cepat jakarta cepat jakarta bandung	Proyek kereta cepat jakarta bandung
paket bijak	Bijak ekonomi paket ekonomi paket bijak ekonomi bijak ekonomi jilid	Paket bijak ekonomi
bank indonesia	Bank indonesia rate Turun suku bunga	Bank indonesia rate
harga minyak	-	Harga Minyak
ekonomi kreatif	Badan ekonomi kreatif	Ekonomi kreatif
rupiah triliun	-	Rupiah trillion
masyarakat ekonomi	Ekonomi asean Masyarakat ekonomi asean Ekonomi asean mea	Masyarakat ekonomi asean
Republika online	-	Republika online
Jusuf kalla	-	Jusuf kalla
Metro news	-	Metro news
Metro news	-	Metro news
Nilai tukar rupiah	-	Nilai tukar rupiah
Indeks harga saham	Harga saham gabung	Indeks harga saham gabungan
Harga daging sapi	-	Harga daging sapi

Jika digambarkan dengan grafik, hasil pengelompokan n-gram adalah seperti pada gambar 4-5.

1. sawit indonesia
2. proyek kereta cepat jakarta bandung <ul style="list-style-type: none"> <li>•kereta cepat</li> <li>•proyek kereta cepat</li> <li>•kereta cepat jakarta</li> <li>•cepat jakarta bandung</li> </ul>
3. paket bijak ekonomi <ul style="list-style-type: none"> <li>•paket bijak</li> <li>•bijak ekonomi</li> <li>•paket bijak ekonomi</li> <li>•bijak ekonomi jilid</li> </ul>
4. bank indonesia rate <ul style="list-style-type: none"> <li>•bank indonesia</li> <li>•bank indoneisa rate</li> <li>•turun suku bunga</li> </ul>
6. harga minyak
7. ekonomi kreatif <ul style="list-style-type: none"> <li>•ekonomi kreatif</li> <li>•badan ekonomi kreatif</li> </ul>
8. rupiah triliun
9. masyarakat ekonomi asean <ul style="list-style-type: none"> <li>•masyarakat ekonomi</li> <li>•ekonomi asean</li> <li>•masyarakat ekonomi asean</li> <li>•ekonomi asean mea</li> </ul>
10. republika online
11. jusef kalla
12. metro news
13. nilai tukarrupiah
14. indeks harga saham gabungan <ul style="list-style-type: none"> <li>•indeks harga saham</li> <li>•harga saham gabung</li> </ul>
harga daging sapi

Gambar 4-5 Pengelompokan N-gram yang Memiliki Kemiripan Makna

Berdasarkan tabel 4.10 dan gambar 4-3, kandidat topik utama yang berhasil diidentifikasi dari kelas kedua ada 14 yaitu :

- |  |                                 |
|--|---------------------------------|
| 1. Sawit indonesia                     | 8. Masyarakat ekonomi asean     |
| 2. Proyek kereta cepat jakarta bandung | 9. Republika online             |
| 3. Paket bijak ekonomi                 | 10. Jusuf kalla                 |
| 4. Bank indonesia rate                 | 11. Metro news                  |
| 5. Harga minyak                        | 12. Nilai tukar rupiah          |
| 6. Ekonomi kreatif                     | 13. Indeks harga saham gabungan |
| 7. Rupiah triliun                      | 14. Harga daging sapi           |

#### 4.7 Identifikasi Jenis Data yang Dihasilkan BPS

Setelah didapatkan 14 kandidat topik utama maka diperlukan penjelasan tentang hubungan kandidat tersebut dengan data yang dihasilkan oleh BPS. Data yang dihasilkan BPS diidentifikasi dari *homepage* resmi BPS di <http://bps.go.id>. Data yang disajikan di *homepage* BPS dibagi menjadi 3 kelompok besar yaitu sosial dan kependudukan, ekonomi dan perdagangan, serta pertanian dan pertambangan. Topik yang sesuai dengan tema penelitian ini adalah kelompok ekonomi dan perdagangan. Hasil pencarian topik data BPS terlihat pada tabel 4.11.

Tabel 4.12 Jenis Data pada Kelompok Ekonomi dan Perdagangan yang Dimiliki oleh Badan Pusat Statistik

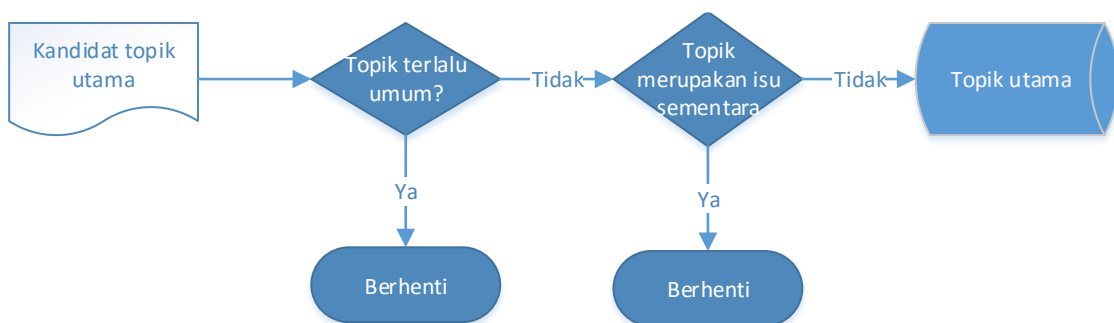
Jenis data	Jenis data
Ekspor-impor	Konstruksi
Energi	Neraca arus dana
Harga eceran	Neraca sosial ekonomi
Harga perdagangan besar	Nilai tukar petani
Harga produsen	Pariwisata
Industri besar dan sedang	Produk domestik bruto (lapangan usaha)
Industri mikro dan kecil	Produk domestik bruto (pengeluaran)

Tabel 4.11 Jenis Data pada Kelompok Ekonomi dan Perdagangan yang Dimiliki oleh Badan Pusat Statistik (lanjutan)

Jenis data	Jenis data
Inflasi	Produk domestik regional bruto (lapangan usaha)
Inflasi	Produk domestik regional bruto (lapangan usaha)
Input-output	Produk domestik regional bruto (pengeluaran)
Indeks tendensi bisnis dan indeks tendensi konsumen	Transportasi
Kuangan	Upah buruh
Komunikasi	Usaha mikro kecil

#### 4.8 Seleksi Kandidat Topik Utama

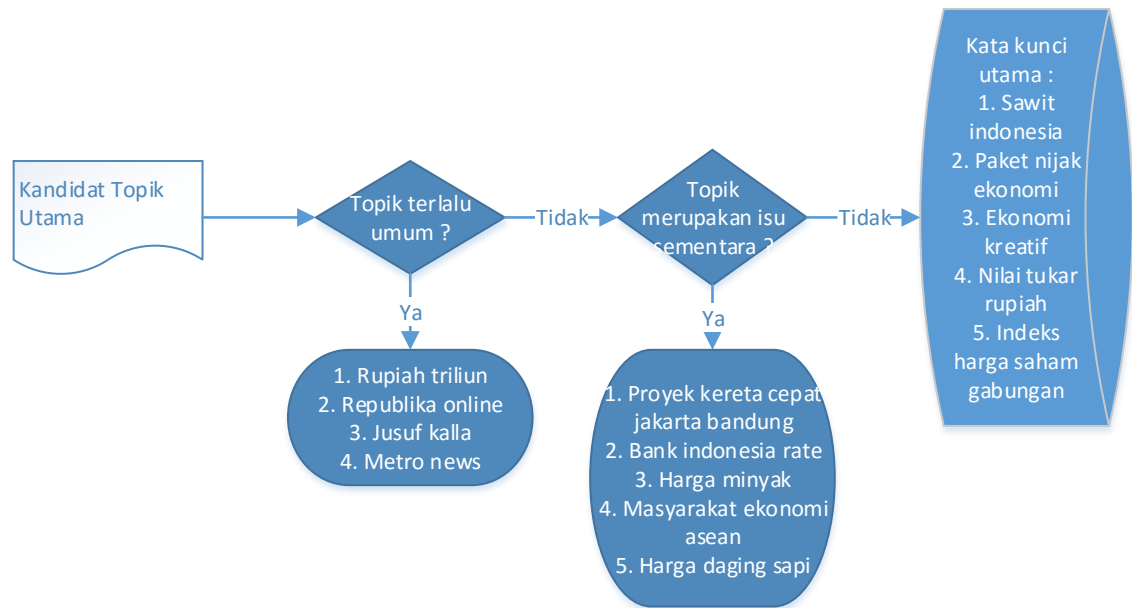
Kandidat topik utama yang berhasil diidentifikasi berjumlah 14 topik. Kandidat topik utama tersebut perlu diperhatikan lebih detail satu per satu. Kandidat topik tersebut masih perlu dilakukan seleksi karena perlu dihubungkan dengan isu yang sedang terjadi pada masa pengumpulan *tweet*. Diagram alir seleksi kandidat topik utama seperti tertera pada gambar 4.6.



Gambar 4-6 Alur Kerja Seleksi Kandidat Topik Utama



Proses seleksi terhadap kandidat topik utama dijelaskan pada gambar 4-7.



Gambar 4-7 Proses Seleksi Kandidat Topik Utama

Pada gambar 4-5 terlihat seleksi dilakukan 2 tahap yaitu melihat apakah kandidat topik utama tersebut terlalu umum untuk dibahas. Pada seleksi pertama terdapat 4 kandidat yang gugur. Kemudian pada seleksi tahap kedua dilakukan pengecekan untuk melihat apakah topik tersebut hanya merupakan isu yang sementara sedang populer. Pada tahap seleksi ini gugur 5 kandidat. Dua tahap seleksi tersebut menghasilkan 5 kandidat yang lolos menjadi topik utama. Penjelasan rinci bisa dilihat pada tabel 4.12.

Tabel 4.13 Hasil Seleksi Kandidat Topik Utama

Kandidat topik utama	Topik terlalu umum	Topik merupakan isu sementara
Sawit indonesia	Tidak	Tidak
Proyek kereta cepat jakarta bandung	Tidak	Ya
Paket bijak ekonomi	Tidak	Tidak
Bank indonesia rate	Tidak	Ya

Tabel 4.12 Hasil Seleksi Kandidat Topik Utama (lanjutan)

<b>Kandidat topik utama</b>	<b>Topik terlalu umum</b>	<b>Topik merupakan isu sementara</b>
Harga minyak	Tidak	Ya
Ekonomi kreatif	Tidak	Tidak
Rupiah triliun	Ya	
Masyarakat ekonomi asean	Tidak	Ya
Republika online	Ya	
Jusuf kalla	Ya	
Metro news	Ya	
Nilai tukar rupiah	Tidak	Tidak
Indeks harga saham gabungan	Tidak	Tidak
Harga daging sapi	Tidak	Ya

Dari tabel 4.12 terlihat bahwa kandidat topik “rupiah triliun”, “republika online”, “jusuf kalla”, dan “metro news” adalah kandidat yang memiliki topik terlalu umum dan tidak relevan dengan pembahasan pada penelitian ini. Kandidat “rupiah triliun” juga tidak memiliki topik yang spesifik. Topik ini bisa digunakan pada ekspor-impor, PRDB, harga, neraca, keuangan, dan topik yang lain. Kandidat “metro news” dan “republika” hanya menunjukkan subyek sumber berita, sehingga kandidat ini tidak spesifik dan relevan. Kandidat topik “jusuf kalla” juga merupakan topik yang terlalu umum mengingat topik tersebut mengacu kepada nama wakil presiden republik indonesia. Keempat topik tersebut dikeluarkan dari kandidat topik utama karena sifatnya yang terlalu umum. Topik tersebut tidak spesifik menyebut sebuah topik sehingga tidak bisa diperkirakan arah pembicaraan mengenai topik tersebut.

Topik “proyek kereta cepat jakarta bandung” diindikasikan menjadi populer karena terdapat beberapa berita kontroversial tentang selama waktu pengumpulan *tweet*. Berita kontroversial tersebut cukup menyita perhatian media

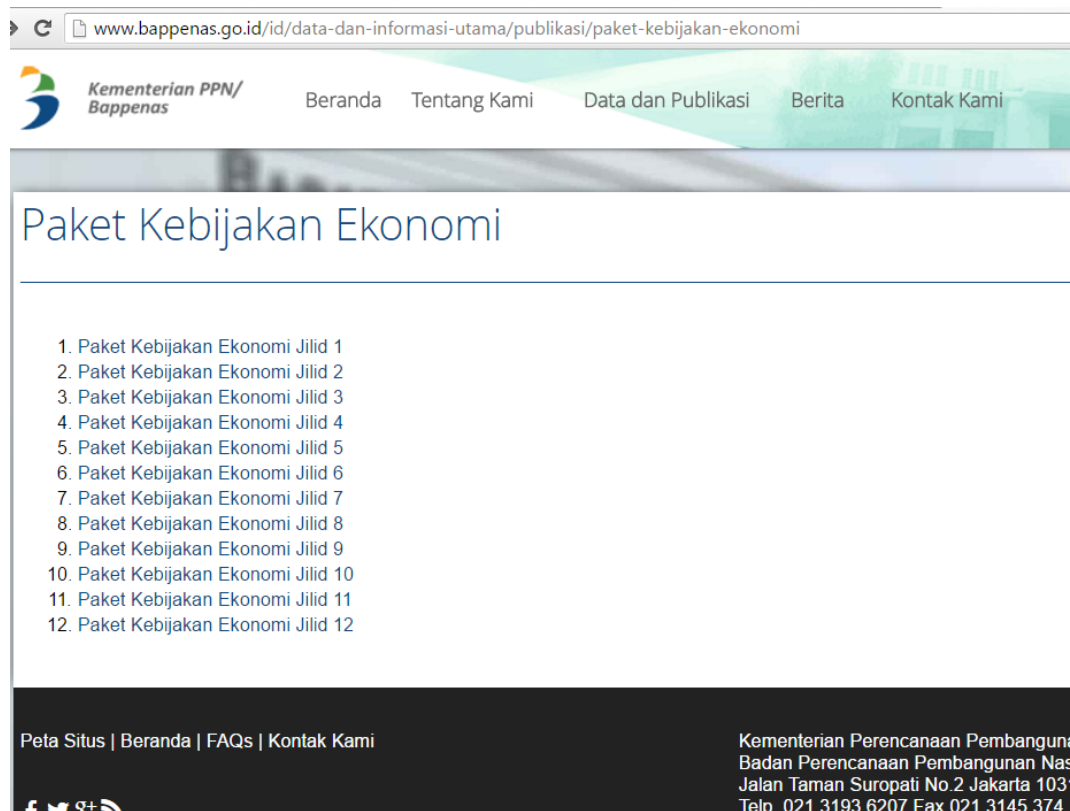
massa sehingga menimbulkan perhatian juga pada masyarakat. Menurut (Hardjanti, R 2016) terdapat beberapa kontroversi mengenai kandidat ini diantaranya :

1. Persaingan jepang dan china pada proyek pembangunan kereta api cepat jakarta-bandung yang akhirnya dimenangkan oleh pihak china. Hal ini menyebabkan pihak jepang merasa kecewa karena merasa lebih layak menang. Bahkan menteri transportasi jepang sempat menyatakan akan meninjau kembali semua hubungan bisnis dengan indonesia.
2. Biaya yang diperlukan proyek ini cukup besar yaitu sekitar 5,5 milyar dolar amerika serikat. Biaya proyek ini murni dilakukan secara business to business yang tidak menggunakan dana APBN. Biaya yang digunakan tersebut berasal dari beberapa BUMN Indonesia dan China. Permasalahan harga juga menjadi polemik ketika diketahui pihak china mengerjakan proyek yang sama di Iran dengan biaya yang jauh lebih murah namun dengan jarak tempuh yang lebih panjang. Hal ini cukup membuat media massa maupun media online menyoroti terus menerus perkembangan proyek ini.
3. Tokoh politik nasional mempermasalahkan urgensi pembangunan proyek ini. Beberapa tokoh partai politik nasional beberapa kali memberikan komentar yang tidak mendukung proyek ini. Seberapa penting proyek ini selalu menjadi masalah utama bagi para tokoh politik nasional. Bahkan ada yang memberikan pendapat bahwa proyek ini tidak sesuai dengan sila keadilan sosial bagi seluruh rakyat indonesia. Kecaman dari beberapa pihak ini dipicu oleh besarnya nilai proyek tersebut.
4. Permasalahan teknis lain seperti ijin pembangunan, ijin amdal, dan beberapa ijin lainnya sempat diberitakan bermasalah.

Dengan adanya beberapa kontroversi pada proyek kereta cepat jakarta-bandung ini topik tersebut dikeluarkan dari kandidat ini dari topik utama. Tingginya kemunculan topik ini kami indikasikan disebabkan oleh gencarnya berita di media massa dan elektronik mengenai kontroversi proyek tersebut pada masa

pengumpulan *tweet*, sehingga kami memperkirakan kandidat ini hanya akan bertahan beberapa waktu saja.

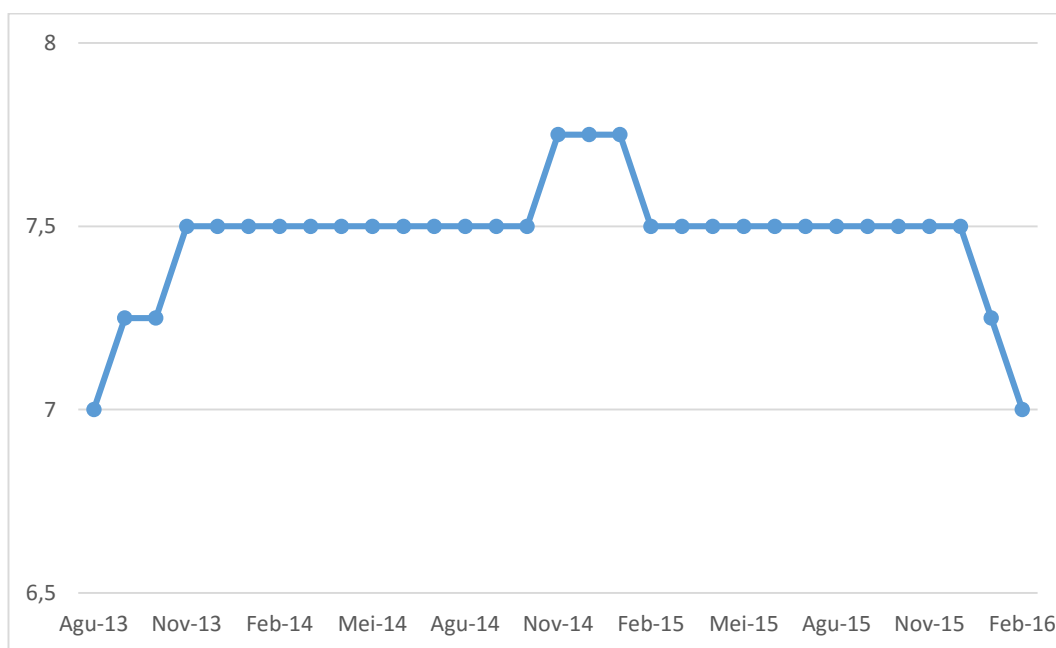
Kandidat topik “paket bijak ekonomi” kami artikan mengacu ke paket kebijakan ekonomi. Topik ini bisa didukung oleh cukup banyak data BPS diantaranya ekspor-impor, pertumbuhan ekonomi (produk domestik bruto), dan tenaga kerja dari BPS. Hal tersebut didukung oleh tingkat kepercayaan presiden republik indonesia terhadap data BPS yang cukup tinggi. Seluruh kebijakan maupun keputusan yang dilakukan oleh presiden republik indonesia menggunakan data BPS sebagai dasar. Namun, dalam proses diseminasi data tersebut tidak disebutkan secara langsung. Topik ini akan lebih bertahan lama karena presiden Republik Indonesia selalu memperbarui paket kebijakan ekonomi jika diperlukan. Hal ini dibuktikan dengan diterbitkannya paket kebijakan ekonomi jilid 10 dan 11 pada masa pengumpulan *tweet*. Gambar 4-8 menunjukkan paket kebijakan ekonomi yang telah dirilis oleh pemerintah.



Gambar 4-8 Rilis Paket Kebijakan Ekonomi Pemerintah Indonesia

Sumber : [www.bappenas.go.id](http://www.bappenas.go.id)

Kandidat topik utama selanjutnya adalah “bank indonesia rate” atau suku bunga bank indonesia. Menurut (BI, 2016) pada bulan Januari dan Februari 2016 terjadi penurunan suku bunga bank indonesia secara berturut-turut dari 7,50% menjadi 7,00%. Penurunan ini bahkan mencapai titik paling rendah dalam kurun waktu 2,5 tahun terakhir. Hal yang jarang terjadi ini bisa menjadi berita yang sering muncul di media massa maupun elektronik. Sehingga pengguna twitter juga terkena imbas populernya berita ini. Gambar 4-9 menunjukkan tren suku bunga bank indonesia.



Gambar 4-9 Tren Suku Bunga Bank Indonesia dari Agustus 2013 sampai dengan Februari 2016

Sumber : [www.bi.go.id](http://www.bi.go.id)

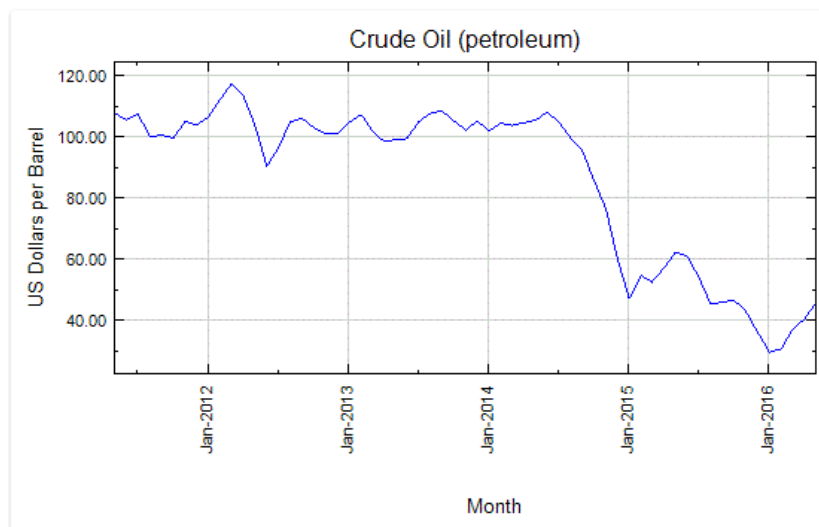
Selain itu, efek yang ditimbulkan dari penurunan suku bunga bank indonesia ini cukup terasa pada kalangan pengusaha dan pebisnis terutama yang berhubungan dengan dunia perbankan. Penurunan suku bunga bank Indonesia ini diindikasikan menjadi penyebab populernya topik ini pada *tweet* yang dikumpulkan. Hal ini diperkirakan akan terjadi sementara waktu saja.

Kandidat topik selanjutnya adalah sawit indonesia. Topik ini terkait dengan data nilai tukar petani, ekspor, industri, dan statistik pertanian dari BPS. Ini adalah salah satu kandidat topik utama pada topik perekonomian yang mewakili

produk pertanian. Tidak ada kasus khusus yang terkait dengan kelapa sawit selama periode pengumpulan *tweet*. Oleh karena itu topik ini digunakan sebagai salah satu topik utama.

Kandidat topik utama selanjutnya yaitu ekonomi kreatif. Topik ini merupakan topik yang cukup baru di Indonesia. Topik ini diperkenalkan oleh Presiden Republik Indonesia mulai tahun 2014. Perwujudan ekonomi kreatif ini diwujudkan dengan pendirian Badan Ekonomi Kreatif di bawah Kementerian Pariwisata. Ekonomi kreatif memiliki pilar utama informasi dan kreativitas. BPS dapat mendukung perkembangan ekonomi kreatif dengan menggunakan data pada bidang komunikasi. Topik ini dimasukkan sebagai topik utama karena diperkirakan topik ini akan berlanjut sesuai dengan tujuan pembangunan oleh presiden republik indonesia.

Kandidat topik selanjutnya adalah harga minyak. Topik ini populer karena tren harga minyak yang terjadi selama periode pengumpulan *tweet*. Pada Januari sampai Februari 2016 ada penurunan harga minyak dunia yang signifikan. Penurunan ini hingga mencapai US \$ 26 per barel yang merupakan harga terendah dalam 13 tahun terakhir. Menurut (Hargaminyak.net, 2016) tren harga minyak tren harga minyak pada tahun 2012 sampai 2016 seperti pada gambar 4-10.



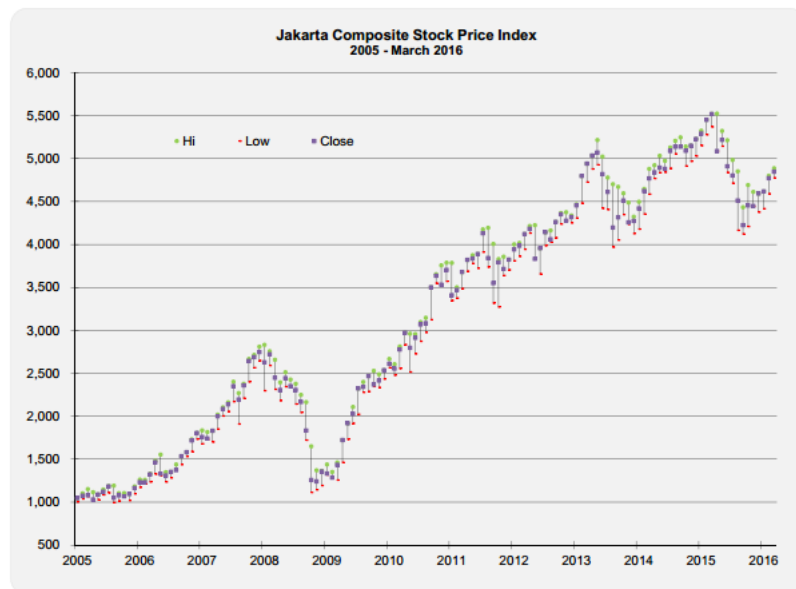
Gambar 4-10 Tren Harga Minyak Dunia 2012-2016

Sumber : [www.hargaminyak.net](http://www.hargaminyak.net)

Kejadian penurunan harga minyak dunia yang ekstrem ini bisa memicu pembahasan dan berita di media massa maupun elektronik. Oleh karena itu popularitas ini hanya dianggap sebagai pengaruh sementara. Sehingga kandidat ini dikeluarkan dari topik utama.

Kandidat topik selanjutnya adalah masyarakat ekonomi asean. Topik ini dapat didukung oleh data BPS di bidang tenaga kerja, pasar dan industri. Menurut (Secretariat. A., 2015) masyarakat ekonomi asean dimulai pada Januari 2016 dengan diberlakukannya pasar bebas antara anggota negara ASEAN. Periode ini sama dengan pengumpulan *tweets*. Oleh karena itu kandidat ini dikeluarkan topik utama.

Kandidat topik berikutnya adalah indeks harga saham. Tidak ada isu khusus pada indeks harga selama periode *tweet* panen. BPS tidak memiliki data mengenai indeks harga saham. Namun, BPS tidak wajib merambah ke topik ini karena data indeks harga saham telah disediakan dengan baik oleh Bursa Efek Indonesia. Oleh karena itu kandidat ini tidak dimasukkan pada kelompok topik utama. Menurut (BEI, 2016) data yang disediakan oleh bursa efek indonesia seperti pada gambar 4-11.

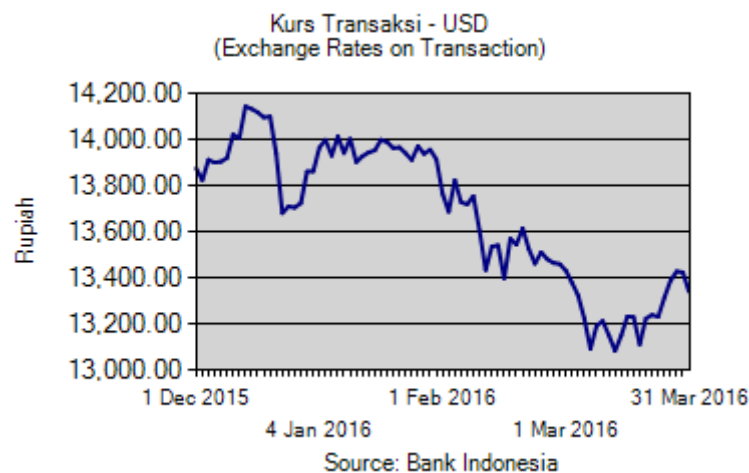


Gambar 4-11 Tren Bursa Efek Indonesia 2005-2016

Sumber : [www.idx.co.id](http://www.idx.co.id)

Kandidat topik selanjutnya adalah harga daging sapi. Pada bulan Januari 2016 terjadi kenaikan yang cukup signifikan pada harga daging sapi. Kenaikan tersebut terjadi di Jakarta. Menurut (Dewi, SM 2016) harga daging sapi yang biasanya pada rentang Rp 80.000 – Rp 90.000 naik menjadi di Rp 120.000. Banyak penjual yang memilih tidak menjual karena adanya kenaikan tersebut. Beberapa portal berita yang membahas kenaikan tersebut. Kenaikan harga ini diindikasikan menyebabkan topik harga daging sapi muncul menjadi kandidat topik utama. Oleh karena itu topik harga daging sapi ini dikeluarkan dari kandidat topik utama.

Kandidat topik terakhir adalah nilai tukar rupiah. Pada periode pengumpulan *tweet* tidak terjadi kenaikan atau perubahan nilai tukar rupiah secara ekstrem. Tidak terdapat juga isu kontroversial mengenai nilai tukar rupiah ini. Tren nilai tukar rupiah terhadap Dollar pada bulan Desember 2015 sampai dengan Maret 2016 terlihat pada gambar 4-12.



Gambar 4-12 Tren Nilai Tukar Rupiah terhadap Dolar Amerika

Sumber : [www.bi.go.id](http://www.bi.go.id)

BPS menyediakan data nilai tukar rupiah setiap satu bulan sekali. Periode terbit data ini dinilai kurang sesuai dengan kebutuhan pasar mengenai data nilai tukar rupiah. Pelaku bisnis pada bidang nilai tukar membutuhkan update data sesering mungkin. Bahkan pergerakan nilai tukar rupiah dalam 1 hari saja menjadi hal yang sangat penting bagi pengusaha dan pebisnis. Data yang terupdate tersebut saat ini banyak tersedia di berbagai media massa maupun elektronik. Bank



indonesia sendiri yang memiliki tugas utama menjaga stabilitas nilai tukar rupiah juga menyediakan data secara *real time* sehingga pengguna data bisa mendapatkan data yang dibutuhkannya.

*Halaman ini sengaja dikosongkan*

## **BAB 5**

### **PENUTUP**

#### **5.1 Kesimpulan**

Dari penelitian yang dilakukan, bisa disimpulkan beberapa hal yaitu :

1. Uji coba pada penelitian ini menunjukkan bahwa model dengan akurasi optimal dicapai menggunakan metode *Decision tree*. Akurasi yang dicapai sebesar 68,03 %. Model tersebut dibangun dengan memberikan batas cacah kata minimal sebesar (*minimum term frequency*) 5 dan *confidence factor* 0,3.
2. Penggunaan unigram pada *tweet* tidak efektif karena kurang menunjukkan topik yang spesifik. Perlu digunakan bigram dan trigram supaya topik utama yang didapatkan menunjukkan topik yang spesifik.
3. Topik-topik utama pada perekonomian yang berhasil diidentifikasi dan diseleksi berjumlah 3 topik, yaitu :
  - a. Paket bijak ekonomi (Paket kebijakan ekonomi)
  - b. Sawit indonesia
  - c. Ekonomi kreatif
4. Strategi diseminasi data BPS menyangkut 3 topik utama tersebut perlu dimaksimalkan. Hal tersebut dikarenakan pada dasarnya BPS memiliki data-data yang mendukung dan/atau berhubungan dengan 3 topik utama tersebut, namun kenyataannya data BPS tidak disebut secara langsung oleh pengguna Twitter.

#### **5.2 Saran**

Penelitian ini perlu dilakukan secara berkelanjutan untuk mengetahui topik yang berkembang, karena topik utama bisa berubah seiring berjalannya waktu. Selain itu juga berfungsi untuk mengetahui konsistensi topik-topik utama yang ditemukan. Jika ada topik utama yang selalu muncul pada periode waktu yang cukup lama maka hal itu akan menjadi informasi yang penting bagi BPS untuk memberikan dukungan data terhadap topik tersebut.

*Halaman ini sengaja dikosongkan*

## DAFTAR PUSTAKA

- APJII, 2015. *Profil Pengguna Internet Indonesia 2014*. 1st ed. Jakarta: APJII.
- BEI, 2016. *Bursa Efek Indonesia*. [Online]  
Available at: [http://www.idx.co.id/Portals/0/StaticData/Publication/Statistic/Quarterly/20160509\\_IDX\\_1st\\_Quarter\\_2016.pdf](http://www.idx.co.id/Portals/0/StaticData/Publication/Statistic/Quarterly/20160509_IDX_1st_Quarter_2016.pdf)  
[Diakses 20 Mei 2016].
- BI, 2016. *Data BI Rate*. [Online]  
Available at: <http://www.bi.go.id/en/moneter/bi-rate/data/Default.aspx>  
[Accessed 12 April 2016].
- BPS, 2011. *Road Map Reformasi Birokrasi Badan Pusat Statistik*. 1st ed. Jakarta: Badan Pusat Statistik.
- Dalyono, C. T., 2010. Media Massa Mempengaruhi Pemanfaatan Teknologi Informasi Generasi Muda dalam Kehidupan Sehari-hari. *Jurnal Ilmu Komunikasi Terakreditasi*, 8(1), pp. 76-86.
- Das, T. K., Acharjya, D. P. & Patra, M. R., 2014. Opinion mining about a product by analyzing public tweets in Twitter. *Computer Communication and Informatics (ICCCI), 2014 International Conference on*, pp. 1-4.
- Dewi, S. N., 2016. *Harga Melonjak, Pedagang Daging Sapi Pilih Tak Jualan*. [Online]  
Available at: <http://bisnis.news.viva.co.id/news/read/727784-harga-melonjak-pedagang-daging-sapi-pilih-tak-jualan>  
[Diakses 4 Februari 2016].
- Hardjanti, R., 2016. *Kereta Cepat Jakarta Bandung, Gebrakan yang Menuai Polemik* : *Okezone Ekonomi*. [Online]  
Available at: <http://economy.okezone.com/read/2016/01/21/320/1293670/kereta-cepat-jakarta-bandung-gebrakan-yang-menuai-polemik>  
[Diakses 20 Mei 2016].
- Hargaminyak.net, 2016. *WTI Crude Oil (Nymex) : Trend Grafik Harga Minyak Dunia*. [Online]  
Available at: <http://hargaminyak.net/trend-grafik/WTI-Crude-Oil-Nymex/>  
[Accessed 23 Mei 2016].
- Kemp, S., 2016. *2016 Digital Yearbook*. [Online]  
Available at: <http://www.slideshare.net/wearesocialsg/2016-digital->

yearbook

[Diakses 7 April 2016].

Khoiron, M., 2016. *Analisa Microblogging Untuk Menentukan Prioritas Kebijakan Publik Berdasarkan Opini Masyarakat Menggunakan Algoritma Naive Bayes Dan Analytical Hierarchy Process* (Doctoral dissertation, Institut Teknologi Sepuluh Nopember).

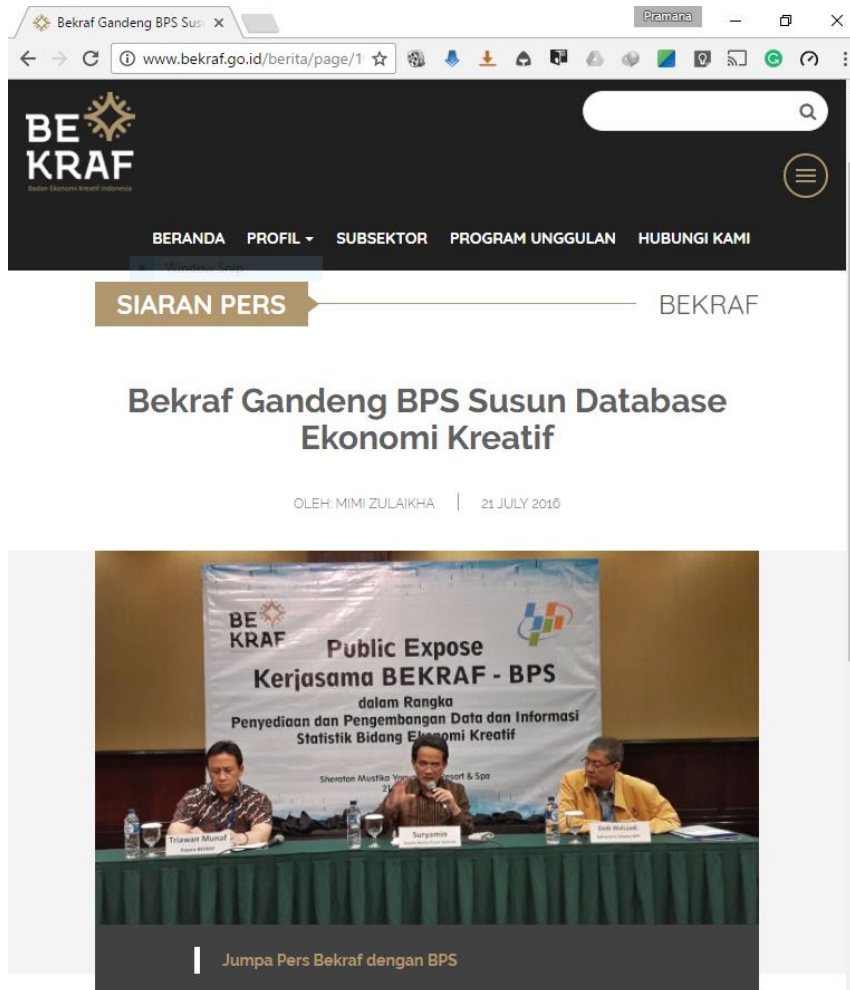
Secretariat, A., 2015. *ASEAN Economic Community Blueprint 2025*. Jakarta: ASEAN Secretariat.

Sembiring, S., 2013. *Kominfo : Pengguna Internet di Indonesia 63 Juta Orang*. [Online]  
Available at: [https://www.kominfo.go.id/content/detail/3415/kominfo-pengguna-internet-di-indonesia-63-juta-orang/0/berita\\_satker](https://www.kominfo.go.id/content/detail/3415/kominfo-pengguna-internet-di-indonesia-63-juta-orang/0/berita_satker)  
[Accessed 7 April 2016].

Tala, F.Z., 2003. *A Study of Stemming Effect on Information Retrieval in Bahasa Indonesia*. Tesis, University of Amsterdam, Amsterdam.

## LAMPIRAN

### Lampiran 1. Kerjasama BEKRAF dengan BPS



SIARAN PERS  
Nomor: 24/SP/BHKP/BEKRAF/VII/2016

#### Bekraf Gandeng BPS Susun Database Ekonomi Kreatif

Yogyakarta, 21 Juli 2016 – Badan Ekonomi Kreatif (Bekraf) bekerjasama dengan Badan Pusat Statistik (BPS) untuk melaksanakan Survei Khusus Ekonomi Kreatif di 57 Kabupaten/Kota pada 34 Provinsi di Indonesia. Kerjasama ini dibentuk dalam rangka usaha Bekraf dalam menyusun database ekonomi kreatif. Survei tersebut ditargetkan mampu menghasilkan data indikator ekonomi berupa Produk Domestik Bruto (PDB), tenaga kerja, dan Ekspor dari ekonomi kreatif dan direktori pelaku usaha ekonomi kreatif di seluruh Indonesia.

“Bekraf bekerjasama dengan BPS dalam mengumpulkan data ekonomi kreatif sebagai upaya untuk melihat peran dan kontribusi ekonomi kreatif terhadap perekonomian Indonesia. Dengan data yang telah terkumpul, kita bisa menentukan arah kebijakan yang tepat bagi perkembangan ekonomi kreatif nasional. Sehingga, target pertumbuhan 12 persen PDB ekonomi kreatif, 13 juta tenaga kerja sektor ekraf, dan kontribusi 10% ekspor ekraf terhadap ekspor nasional hingga tahun 2019 dapat tercapai,” ungkap Kepala Bekraf Triawan Munaf.

Kerjasama ini dilatarbelakangi oleh minimnya data ekonomi kreatif Indonesia yang

*Halaman ini sengaja dikosongkan*



## BIOGRAFI PENULIS



**Pramana Yhoga Chandra Kusuma**, Magister di Institut Teknologi Sepuluh Nopember, Fakultas Teknologi Industri, Jurusan Teknik Elektro, Bidang Keahlian Telematika Chief Information Officer (CIO), merupakan salah satu penerima kerjasama Kementerian Komunikasi dan Informatika Republik Indonesia dengan Badan Pusat Statistik Republik Indonesia. Lahir di Tulungagung pada tanggal 7 September 1984. Anak pertama dari dua bersaudara dari pasangan Tri Purnomo dan Amirulsari. Sudah menikah dengan Reni Puspitasari dan dikaruniai 1 orang putra dan 1 orang putri bernama Muhammad Fahri Kamil dan Aisyah Kalyani Prameswari. Saat menjalankan pendidikan, penulis bertugas di Badan Pusat Statistik Provinsi Gorontalo.

Alamat email : [pramanayhoga@gmail.com](mailto:pramanayhoga@gmail.com)

### **Riwayat Pendidikan :**

**SD** : SDN Kiping 2, Kec. Gondang, Tulungagung

**SMP** : SMPN 1 Gondang, Kec. Gondang, Tulungagung

**SMU** : SMU 1 Boyolangu, Tulungagung

**D4** : Sekolah Tinggi Ilmu Statistik (STIS)

### **Riwayat Pekerjaan :**

1. Staf Seksi Statistik Produksi BPS Boalemo, Gorontalo
2. Staf IPDS BPS Boalemo, Gorontalo
3. Kepala Seksi IPDS BPS Boalemo, Gorontalo
4. Kepala Seksi Jaringan dan Rujukan Statistik BPS Provinsi Gorontalo

*Halaman ini sengaja dikosongkan*